ENDOMANNOSIDASES IN THE MODIFICATION OF GLYCOPROTEINS IN EUKARYOTES

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation-in-part of U. S. Application No. 10/371,877, filed on February 20, 2003.

FIELD OF THE INVENTION

[0002] The present invention generally relates to methods of modifying the glycosylation structures of recombinant proteins expressed in fungi or other lower eukaryotes, to more closely resemble the glycosylation of proteins from higher mammals, in particular humans. The present invention also relates to novel enzymes and, nucleic acids encoding them and, hosts engineered to express the enzymes, methods for producing modified glycoproteins in hosts and modified glycoproteins so produced.

BACKGROUND OF THE INVENTION

[0003] After DNA is transcribed and translated into a protein, further post-translational processing involves the attachment of sugar residues, a process known as glycosylation. Different organisms produce different glycosylation enzymes (glycosyltransferases and glycosidases) and have different substrates (nucleotide sugars) available, so that the glycosylation patterns as well as composition of the individual oligosaccharides, even of one and the same protein, will be different depending on the host system in which the particular protein is being expressed.

Express Mail Label No. EV132184215US

Bacteria typically do not glycosylate proteins and if so only in a very unspecific manner (Moens and Vanderleyden, *Arch. Microbiol.* 168(3):169-175 (1997)). Lower eukaryotes such as filamentous fungi and yeast add primarily mannose and mannosylphosphate sugars, whereas insect cells such as Sf9 cells glycosylate proteins in yet another way. See R.K. Bretthauer et al., *Biotechnology and Applied Biochemistry* 1999 30:193-200 (1999); W. Martinet, et al., *Biotechnology Letters* 1998 20:1171-1177 (1998); S. Weikert, et al., *Nature Biotechnology* 1999 17: 1116-1121 (1999); M. Malissard, et al., *Biochem.Biophys.Res.Comm.* 2000 267:169-173 (2000); D. Jarvis, et al., *Curr. Op. Biotech.* 1998 9:528-533 (1998); and Takeuchi, *Trends in Glycoscience and Glycotechnology* 1997 9:S29-S35 (1997).

[0004] N-linked glycosylation plays a major role in the processing of many cellular and secreted proteins. In eukaryotes, the preassembled oligosaccharide Glc3Man9GlcNAc2 is transferred from dolichol onto the acceptor site of the protein by oligosaccharyltransferase in the endoplasmic reticulum (Dempski and Imperiali, Curr. Opin. Chem. Biol. 6: 844-850 (2002)). Subsequently, the terminal a-1,2-glucose is removed by glucosidase I facilitating the removal of the remaining two a-1,3-glucose residues by glucosidase II (Herscovics, Biochim. Biophys. Acta 1473: 96-107 (1999)). The high mannose glycan remaining is processed by the ER mannosidase, to Man8GlcNAc2, prior to translocation of the glycoprotein to the Golgi, where the glycan structure is further modified. Incorrect processing of the glycan structure in the ER, in turn, can prevent subsequent modification, leading to a disease state. The absence of glucosidase I results in congenital disorder of glycosylation type (CDG) IIb which is extremely rare, with only one reported human case, and leads to early death (Marquardt and Denecke, *Eur. J. Pediatr.* 162: 359-379 (2003)). Isolation of the Chinese hamster ovary cell line Lec23, deficient in glucosidase I, demonstrated that the predominant glycoform present is Glc3Man9GlcNAc2 (Ray et al., *J. Biol. Chem.* 266: 22818-22825 (1991)). The initial stages of glycosylation in yeast and mammals are identical with the same glycan structures emerging from the endoplasmic reticulum. However, when these glycans are processed by the Golgi, the resultant structures are drastically different, thus resulting in yeast glycosylation patterns that differ substantially

from those found in higher eukaryotes, such as humans and other mammals (R. Bretthauer, et al., *Biotechnology and Applied Biochemistry* 30:193-200 (1999)). Moreover, the vastly different glycosylation pattern has, in some cases, been shown to increase the immunogenicity of these proteins in humans and reduce their half-life (Takeuchi (1997) *supra*).

[0005] The early steps of human glycosylation can be divided into at least two different phases: (i) lipid-linked Glc3Man9GlcNAc2 oligosaccharides assembled by a sequential set of reactions at the membrane of the endoplasmatic reticulum (ER); and (ii) the transfer of this oligosaccharide from the lipid anchor dolichyl pyrophosphate on to de novo synthesized protein. The site of the specific transfer is defined by an Asparagine residue in the sequence Asn-Xaa-Ser/Thr, where Xaa can be any amino acid except Proline (Y. Gavel et al., *Protein Engineering* 3:433-442 (1990)).

[0006] Further processing by glucosidases and mannosidases occurs in the ER before the nascent glycoprotein is transferred to the early Golgi apparatus, where additional mannose residues are removed by Golgi specific a-1,2-mannosidases. Processing continues as the protein proceeds through the Golgi. In the medial Golgi, a number of modifying enzymes, including N-acetylglucosaminyl-transferases (GnT I, GnT II, GnT III, GnT IV GnT V GnT VI), mannosidase II, and fucosyltransferases, add and remove specific sugar residues. Finally, in the trans-Golgi, galactosyltranferases and sialyltransferases produce a structure that is released from the Golgi. The glycans characterized as bi-, tri- and tetra-antennary structures containing galactose, fucose, N-acetylglucosamine and a high degree of terminal sialic acid give glycoproteins their human characteristics.

[0007] When proteins are isolated from humans or animals, a significant number of them are post-translationally modified, with glycosylation being one of the most significant modifications. Several studies have shown that glycosylation plays an important role in determining the (1) immunogenicity, (2) pharmacokinetic properties, (3) trafficking, and (4) efficacy of therapeutic proteins. An estimated 70% of all therapeutic proteins are glycosylated and thus currently rely on a production system (i.e., host) that is able to glycosylate in a manner similar to humans. To date, most glycoproteins are made in a mammalian host system. It is

thus not surprising that substantial efforts by the pharmaceutical industry have been directed at developing processes to obtain glycoproteins that are as "humanoid" as possible. This may involve the genetic engineering of such mammalian cells to enhance the degree of sialylation (i.e., terminal addition of sialic acid) of proteins expressed by the cells, which is known to improve pharmacokinetic properties of such proteins. Alternatively, one may improve the degree of sialylation by in vitro addition of such sugars by using known glycosyltransferases and their respective nucleotide sugar substrates (e.g. 2,3 sialyltransferase and CMP-Sialic acid).

[0008] Further research may reveal the biological and therapeutic significance of specific glycoforms, thereby rendering the ability to produce such specific glycoforms desirable. To date, efforts have concentrated on making proteins with fairly well characterized glycosylation patterns, and expressing a cDNA encoding such a protein in one of the following higher eukaryotic protein expression systems:

1. Higher eukaryotes such as Chinese hamster ovary cells (CHO), mouse fibroblast cells and mouse myeloma cells (R. Werner, et al., Arzneimittel-Forschung-Drug rResearch 1998 48:870-880 (1998));

2. Transgenic animals such as goats, sheep, mice and others (Dente et al., *Genes and Development* 2:259-266 (1988); Cole et al., *J. Cell. Biochem.* 265:supplement 18D (1994); P. McGarvey et al., *Biotechnology* 13:1484-1487 (1995); Bardor et al., *Trends in Plant Science* 4:376-380 (1999));

3. Plants (Arabidopsis thaliana, tobacco etc.) (Staub et al., *Nature Biotechnology* 18:333-338 (2000); McGarvey et al., *Biotechnology* 13:1484-1487 (1995); Bardor et al., *Trends in Plant Science* 4:376-380 (1999));

4. Insect cells (Spodoptera frugiperda Sf9, Sf21, Trichoplusia ni, etc. in combination with recombinant baculorviruses such as Autographa californica multiple nuclear polyhedrosis virus which infects lepidopteran cells (Altmann, et al., *Glycoconjugate Journal* 16:109-123 (1999)).

[0009] While most higher eukaryotes carry out glycosylation reactions that are similar to those found in humans, recombinant human proteins expressed in the above mentioned host systems invariably differ from their "natural" human

counterpart (Raju, et al. *Glycobiology* 10:477-486 (2000)). Extensive development work has thus been directed at finding ways to improving the "human character" of proteins made in these expression systems. This includes the optimization of fermentation conditions and the genetic modification of protein expression hosts by introducing genes encoding enzymes involved in the formation of human like glycoforms (Werner et al., *Arzneimittel-Forschung-Drug Res.* 48:870-880 (1998); Weikert et al. *Nature Biotechnology* 17:1116-1121 (1999); Andersen et al., *Current Opinion in Biotechnology* 5:546-549 (1994); Yang et al., *Biotechnology and Bioengineering* 68:370-380 (2000)).

[0010] What has not been solved, however, are the inherent problems associated with all mammalian expression systems. Fermentation processes based on mammalian cell culture (e.g. CHO, Murine, or more recently, human cells) tend to be very slow (fermentation times in excess of one week are not uncommon), often yield low product titers, require expensive nutrients and cofactors (e.g. bovine fetal serum), are limited by programmed cell death (apoptosis), and often do not allow for the expression of particular therapeutically valuable proteins. More importantly, mammalian cells are susceptible to viruses that have the potential to be human pathogens and stringent quality controls are required to assure product safety. This is of particular concern since as many such processes require the addition of complex and temperature sensitive media components that are derived from animals (e.g. bovine calf serum), which may carry agents pathogenic to humans such as bovine spongiform encephalopathy (BSE) prions or viruses.

[0011] The production of therapeutic compounds is preferably carried out in a well-controlled sterile environment. An animal farm, no matter how cleanly kept, does not constitute such an environment. Transgenic animals are currently considered for manufacturing high volume therapeutic proteins such as: human serum albumin, tissue plasminogen activator, monoclonal antibodies, hemoglobin, collagen, fibrinogen and others. While transgenic goats and other transgenic animals (mice, sheep, cows, etc.) can be genetically engineered to produce therapeutic proteins at high concentrations in the milk, recovery is burdensome since every batch has to undergo rigorous quality control. A transgenic goat may produce sufficient quantities of a therapeutic protein over the course of a year,

however, every batch of milk has to be inspected and checked for contamination by bacteria, fungi, viruses and prions. This requires an extensive quality control and assurance infrastructure to ensure product safety and regulatory compliance. In the case of scrapies and bovine spongiform encephalopathy, testing can take about a year to rule out infection. In the interim, trust in a reliable source of animals substitutes for an actual proof of absence. Whereas cells grown in a fermenter are derived from one well characterized Master Cell Bank (MCB), transgenic technology relies on different animals and thus is inherently non-uniform. Furthermore, external factors such as different food uptake, disease and lack of homogeneity within a herd may affect glycosylation patterns of the final product. It is known in humans, for example, that different dietary habits impact glycosylation patterns, and it is thus prudent to expect a similar effect in animals. Producing the same protein in fewer batch fermentations would be (1) more practical, (2) safer, and (3) cheaper, and thus preferable.

[0012] Transgenic plants have emerged as a potential source to obtain proteins of therapeutic value. However, high level expression of proteins in plants suffers from gene silencing, a mechanism by which highly expressed proteins are down regulated in subsequent generations. In addition, it is known that plants add xylose and a-1,3 linked fucose, a glycosylation pattern that is usually not found in human glycoproteins, and has shown to lead to immunogenic side effects in higher mammals. Growing transgenic plants in an open field does not constitute a well-controlled production environment. Recovery of proteins from plants is not a trivial matter and has yet to demonstrate cost competitiveness with the recovery of secreted proteins in a fermenter.

[0013] Most currently produced therapeutic glycoproteins are therefore expressed in mammalian cells and much effort has been directed at improving (i.e., humanizing) the glycosylation pattern of these recombinant proteins. Changes in medium composition as well as the co-expression of genes encoding enzymes involved in human glycosylation have been successfully employed (see, for example, Weikert et al., *Nature Biotechnology* 17:1116-1121 (1999)).

[0014] While recombinant proteins similar to their human counterparts can be made in mammalian expression systems, it is currently not possible to make

proteins with a humanoid glycosylation pattern in lower eukaryotes (e.g., fungi and yeast). Although the core oligosaccharide structure transferred to the protein in the endoplasmic reticulum is basically identical in mammals and lower eukaryotes, substantial differences have been found in the subsequent processing reactions of the Golgi apparatus of fungi and mammals. In fact, even amongst different lower eukaryotes, there exists a great variety of glycosylation structures. This has prevented the use of lower eukaryotes as hosts for the production of recombinant human glycoproteins despite otherwise notable advantages over mammalian expression systems, such as: (1) generally higher product titers, (2) shorter fermentation times, (3) having an alternative for proteins that are poorly expressed in mammalian cells, (4) the ability to grow in a chemically defined protein free medium and thus not requiring complex animal derived media components, and (5) and the absence of retroviral infections of such hosts.

[0015] Various methylotrophic yeasts such as Pichia pastoris, Pichia methanolica, and Hansenula polymorpha, have played particularly important roles as eukaryotic expression systems since because they are able to grow to high cell densities and secrete large quantities of recombinant protein. However, as noted above, lower eukaryotes such as yeast do not glycosylate proteins like higher mammals. See, for example, U.S. Patent No. 5,834,251 to Maras et al. (1994). Maras and Contreras have shown recently that P. pastoris is not inherently able to produce useful quantities (greater than 5%) of GlcNAcTransferase I accepting carbohydrate. (Martinet et al., *Biotechnology Letters* 20:1171-1177 (1998)). Chiba et al. (*J. Biol. Chem.* 273: 26298-26304 (1998)) have shown that *S. cerevisiae* can be engineered to provide structures ranging from $Man_8GlcNAc_2$ to $Man_5GlcNAc_2$ structures, by eliminating 1,6 mannosyltransferase (OCH1), 1,3 mannosyltransferase (MNN1) and mannosylphosphatetransferase (MNN4) and by targeting the catalytic domain of a-1,2-mannosidase I from Aspergillus saitoi into the ER of *S. cerevisiae*, by using a ER retrieval/targeting sequence (Chiba 1998, *supra*). However, this attempt resulted in little or no production of the desired $Man_5GlcNAc_2$. The model protein (carboxypeptidase Y) was trimmed to give a mixture consisting of 27% $Man_5GlcNAc_2$, 22% $Man_6GlcNAc_2$, 22% $Man_7GlcNAc_2$, 29% $Man_8GlcNAc_2$. As only the $Man_5GlcNAc_2$ glycans are

susceptible to further enzymatic conversion to human glycoforms, this approach is very inefficient for the following reasons: In proteins having a single N-glycosylation site, at least 73% of all N-glycans will not be available for modification by GlcNAc transferase I. In a protein having two or three N-glycosylation sites, at least 93% or 98%, respectively, would not be accessible for modification by GlcNAc transferase I. Such low efficiencies of conversion are unsatisfactory for the production of therapeutic agents; given the large number of modifying steps each cloned enzyme needs to function at highest possible efficiency.

[0016] A number of reasons may explain the inefficiency in the production of glycan formation mentioned above. This may, in part, be due to the inefficient processing of glycans in the ER either by glucosidase I, II or resident ER mannosidase. A recently evolved class of mannosidase proteins has been identified in eukaryotes of the chordate phylum (including mammals, birds, reptiles, amphibians and fish) that is also involved in glucose removal. These glycosidic enzymes have been defined as endomannosidases. The activity of the endomannosidases has been characterized in the processing of N-linked oligosaccharides, namely, in removing a glucose α1,3 mannose dissacharide. The utility in removing of the glucose and mannose residues on oligosaccharides in the initial steps of N-linked oligosaccharide processing is known to be useful for the production of complex carbohydrates has been well-established.

Although endomannosidases were originally detected in the trimming of $GlcMan_9GlcNAc_2$ to $Man_8GlcNAc_2$, they also process other glucosylated structures (Fig. 1). Overall, mono-glucosylated glycans are most efficiently modified although di- and tri-glucosylated glycans may also be processed to a lesser extent (Lubas et al., *J. Biol. Chem.* 263(8):3990-8 (1988)). Furthermore, not only is $GlcMan_9GlcNAc_2$ is the preferred substrate but other monoglucosylated glycans, such as $GlcMan_7GlcNAc_2$ and $GlcMan_5GlcNAc_2$, are trimmed (to $Man_6GlcNAc_2$ and $Man_4GlcNAc_2$, respectively) just as efficiently. The occurrence of this class of proteins so late in evolution suggests that this is a unique requirement to enhance the pronounced trimming of N-linked glycans, as observed in higher eukaryotes. This suggestion is further strengthened by the fact that

endomannosidase is located in the Golgi and not the ER where complete deglucosylation has traditionally been reported to occur.

[0017] Previous research has shown that glucose excision occurs primarily in the ER through sequential action of glucosidase I and II (Moremen et al., *Glycobiology* 4: 113-125 (1994)). However, more recent research suggests the apparent alternate glucosidase II – independent deglucosylation pathway involving a quality control mechanism in the Golgi apparatus (Zuber et al., *Mol. Biol. Cell.* Dec;11(12): 4227-40 (2000)). Studies in glucosidase II- deficient mouse lymphoma cells show evidence of the deglucosylation mechanism by the endomannosidase (Moore et al., *J. Biol. Chem.* 267(12):8443-51 (1992)). Furthermore, a mouse lymphoma cell line, PHAR2.7, has been isolated which has no glucosidase II activity resulting primarily in the production of the glycoforms $Glc_2Man_9GlcNAc_2$ and $Glc_2Man_8GlcNAc_2$ (Reitman et al., *J. Biol. Chem.* 257: 10357-10363 (1982)). Analysis of this latter cell line demonstrated that, despite the absence of glucosidase II, deglucosylated high mannose structures were present, thus, indicating the existence of an alternative processing pathway for glucosylated structures (Moore and Spiro, *J. Biol. Chem.* 267: 8443-8451 (1992)). The enzyme responsible for this glucosidase-independent pathway has been identified as endomannosidase (E.C. 3.2.1.130). Endomannosidase catalyzes the hydrolysis of mono-, di- and tri-glucosylated high mannose glycoforms, removing the glucose residue(s) present and the juxta-positioned mannose (Hiraizumi et al., *J. Biol. Chem.* 268: 9927-9935 (1993); Bause and Burbach, *Biol. Chem.* 377: 639-646 (1996)).

[0018] The endomannosidase does not appear to distinguish between differing mannose structures of a glucosylated glycoform, hydrolyzing $Glc_1Man_{9\text{-}5}GlcNAc_2$ to $Man_{8\text{-}4}GlcNAc_2$ (Lubas and Spiro, *J. Biol. Chem.* 263: 3990-3998 (1988)). To date, the only endomannosidase to have been cloned is from the rat liver. Rat liver endomannosidase encodes a predicted open reading frame (ORF) of 451 amino acids with a molecular mass of 52 kDa (Spiro et al., *J. Biol. Chem.* 272: 29356-29363 (1997)). This enzyme has a neutral pH optimum and does not appear to have any specific cation requirement (Bause and Burbach 1996, *supra*). Unlike the glucosidase enzymes, which are localized in the ER, the endomannosidase is

primarily localized in the Golgi (Zuber et al., *Mol. Biol. Cell* 11: 4227-4240 (2000)), suggesting that it may play a quality control role by processing glucosylated glycoforms leaking from the ER.

[0019] Given the utility of modifying glucosylated glycans for the production of human-like glycoproteins, a method for modifying glucosylated glycans by expressing an endomannosidase activity in a host cell would be desirable.

SUMMARY OF THE INVENTION

[0020] Methods have been developed for modifying a glucosylated N-glycan by genetically engineering strains of non-mammalian eukaryotes which are able to produce recombinant glycoproteins substantially equivalent to their human counterparts. These cell lines, including yeast, filamentous fungi, insect cells, and plant cells grown in suspension culture, have genetically modified glycosylation pathways allowing them to carry out a sequence of enzymatic reactions which mimic the processing of glycoproteins in humans. As described herein, strains have been developed to express catalytically active endomannosidase genes to enhance the processing of the N-linked glycan structures with the overall goal of obtaining a more human-like glycan structure. In addition, cloning and expression of a novel human and mouse endomannosidase are also disclosed. The method of the present invention can be adapted to engineer cell lines having desired glycosylation structures useful in the production of therapeutic proteins.

BRIEF DESCRIPTION OF THE DRAWINGS

[0021] **Fig. 1** is a schematic diagram of an endomannosidase modifying mono-, di- and tri- glucosylated glycans in the Golgi in comparison to glucose processing of N-glycans in the ER. Highlighted are additional glucose residues that can be hydrolyzed.

[0022] **Fig. 2** is a schematic diagram of an endomannosidase processing the glucosylated structure $Glc_3Man_9GlcNAc_2$ to $Man_5GlcNAc_2$ glycans in the Golgi. Highlighted mannose residues represent constituents which, in various combinations, produce various types of high mannan glycans that may be substrates for the endomannosidase.

[0023] **Fig. 3** shows a BLAST analysis of rat endomannosidase to identify homologues. Panel **A** shows identification of a human sequence showing 88% identity to the C-terminus of rat endomannosidase. Panel **B** shows the N-terminus of isolated sequence from Panel **A** which was used to isolate the 5' region of the human endomannosidase in Panel **C**. Panel **C** shows sequence of the potential N-terminus of human endomannosidase.

[0024] **Fig. 4** shows nucleotide and amino acid sequences of human liver endomannosidase. Nucleotide sequence (*upper*) and one-letter amino acid sequence (*lower*) of human endomannosidase are shown with residue numbers labeled on the left. The nucleotide region in bold represents the overlapping segments of Genbank sequences gi:18031878 (*underlined*) and gi:20547442 (*regular text*) used to assemble the putative full-length human liver endomannosidase. The putative transmembrane domain identified by Kyte and Doolittle analysis (*J. Mol. Biol.* 157: 105-132 (1982)) (see **Fig. 5**) is highlighted by an open box.

[0025] **Fig. 5** shows the hydropathy plot of the amino acid sequence of the human endomannosidase, produced according to the method of Kyte and Doolittle ((1982) *supra*), using the web-based software GREASE and a window of 11 residues. The *filled-in box* represents an N-terminal region of high hydrophobicity, suggesting the presence of a putative transmembrane domain. This region is also represented in **Fig. 4** by an open box (amino acid residues 10-26).

[0026] **Fig. 6** shows nucleotide and amino acid sequences of mouse endomannosidase (Genbank AK030141). Nucleotide sequence (*upper*) and one-letter amino acid sequence (*lower*) of mouse endomannosidase are shown with residue numbers labeled on the left. The putative transmembrane domain identified by Kyte and Doolittle analysis (*J. Mol. Biol.* 157: 105-132 (1982)) is highlighted by an open box.

[0027] **Fig. 7** shows the alignment of three endomannosidase open-reading frames. The human, mouse and rat endomannosidase ORFs were aligned using the Megalign software of the DNASTAR suite of programs. The algorithm chosen for the analysis was the CLUSTAL V version (Higgins and Sharp *Comput. Appl. Biosci.* 5, 151-153 (1989)). Residues displayed by shading represent amino acids

that are identical between at least two of the ORFs. The amino acid position of each ORF is presented to the left of the aligned sequence.

[0028] **Fig. 8** depicts a Northern blot analysis of RNAs from a variety of human tissues hybridized with a labeled human endomannosidase nucleic acid probe.

[0029] **Fig. 9** depicts a Western blot analysis of prepurification on Ni-resin of secreted N-terminal tagged endomannosidase, samples from control (GS115) (**A**), rEndo (**YSH89**) (**B**) and hEndo (**YSH90**) (**C**) strains. The samples were detected using anti-FLAG M2 antibody (Stratagene, La Jolla, CA).

[0030] **Fig. 10A** shows a MALDI-TOF MS analysis of N-glycans isolated from a kringle 3 glycoprotein produced in *P.pastoris* **RDP-25** (*och1 alg3*).

[0031] **Fig. 10B** shows a MALDI-TOF MS analysis of N-glycans isolated from a kringle 3 glycoprotein produced in *P.pastoris* **RDP-25** (*och1 alg3*) transformed with **pSH280** (rat endomannosidaseΔ48/*Mnn*11(m)) showing, a peak, among others, at 1099 m/z **[c]** corresponding to the mass of $Man_4GlcNAc_2$ and 1424 m/z **[a]** corresponding to the mass of hexose 6. This strain was designated as **YSH97**.

[0032] **Fig. 10C** shows a MALDI-TOF MS analysis of N-glycans isolated from a kringle 3 glycoprotein produced in *P.pastoris* **YSH97** after *in vitro* digestion with α1,2-mannosidase, exhibiting a peak at 938 m/z **[b]** (Na^+ adduct) corresponding to the mass of $Man_3GlcNAc_2$.

[0033] **Fig. 11A** shows a MALDI-TOF MS analysis of N-glycans isolated from a kringle 3 glycoprotein produced in *P.pastoris* **RDP-25** (*och1 alg3*).

[0034] **Fig. 11B** shows a MALDI-TOF MS analysis of N-glycans isolated from a kringle 3 glycoprotein produced in *P.pastoris* **RDP-25** (*och1 alg3*) transformed with **pSH279** (rat endomannosidaseΔ48/*Van*1(s)) showing among others, a peak at 1116 m/z **[c]** corresponding to the mass of $Man_4GlcNAc_2$ and 1441 m/z **[a]** corresponding to the mass of hexose 6. This strain was designated **YSH96**.

[0035] **Fig. 11C** shows a MALDI-TOF MS analysis of N-glycans isolated from a kringle 3 glycoprotein produced in *P.pastoris* **YSH96** after *in vitro* digestion with α1,2-mannosidase, exhibiting a peak at 938 m/z **[b]** (Na^+ adduct) corresponding to the mass of $Man_3GlcNAc_2$ and a second peak at 1425 m/z **[a]** showing a decrease in hexose 6.

[0036] **Fig. 12A** shows a MALDI-TOF MS analysis of N-glycans isolated from a kringle 3 glycoprotein produced in *P.pastoris* **RDP-25** (*och1 alg3*).

[0037] **Fig. 12B** shows a MALDI-TOF MS analysis of N-glycans isolated from a kringle 3 glycoprotein produced in *P.pastoris* **RDP-25** (*och1 alg3*) transformed with **pSH278** (rat endomannosidaseΔ48/*Gls*1(s)) showing, a peak, among others, at 1439 m/z (K^+ adduct) **[c]** and a peak at 1422 m/z (Na^+ adduct) corresponding to the mass of hexose 6 **[a]**. This strain was designated **YSH95**.

[0038] **Fig. 12C** shows a MALDI-TOF MS analysis of N-glycans isolated from a kringle 3 glycoprotein produced in *P.pastoris* **YSH95** after *in vitro* digestion with α1,2-mannosidase, exhibiting a peak at 936 m/z **[b]** (Na^+ adduct) corresponding to the mass of $Man_3GlcNAc_2$ and a peak at 1423 m/z **[a]** showing a decrease in hexose 6.

[0039] **Fig. 13** shows a high performance liquid chromatogram *in vitro* assay for rat and human endomannosidase activity. Panel **A** shows the hexose 6 standard $GlcMan_5GlcNAc_2$ in BMMY. Panel **B** shows glycan substrate produced from rat endomannosidase incubated with supernatant from *P. pastoris* **YSH13**. Panel **C** shows glycan substrate produced from human endomannosidase incubated with supernatant from *P. pastoris* **YSH16**. See **Fig. 14** for structures corresponding to (i) and (ii).

[0040] **Fig. 14** represents substrate glycan modification by endomannosidase and subsequent confirmation of product structure by α1,2-mannosidase digestion and analysis. Structures illustrated are $GlcMan_5GlcNAc_2$ **(i)**, $Man_4GlcNAc_2$ **(ii)** and $Man_3GlcNAc_2$ **(iii)**. R represents the reducing terminus of the glycan. The substrate $GlcMan_5GlcNAc_2$ **(i)** is modified by an endomannosidase converting it to $Man_4GlcNAc_2$ **(ii)** (hydrolyzing Glcα1,3Man). Subsequent α1,2-mannosidase digestion results in $Man_3GlcNAc_2$ **(iii)**.

[0041] **Fig. 15** shows a pH profile of the activity of human endomannosidase, indicated as % of $GlcMan_5GlcNAc_2$ substrate converted to $Man_4GlcNAc_2$ as a function of pH.

DETAILED DESCRIPTION OF THE INVENTION

[0042] Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly

understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. The methods and techniques of the present invention are generally performed according to conventional methods well known in the art. Generally, nomenclatures used in connection with, and techniques of biochemistry, enzymology, molecular and cellular biology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well known and commonly used in the art. The methods and techniques of the present invention are generally performed according to conventional methods well known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. See, e.g., Sambrook et al. Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989); Ausubel et al., Current Protocols in Molecular Biology, Greene Publishing Associates (1992, and Supplements to 2002); Harlow and Lane Antibodies: A Laboratory Manual Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1990); Introduction to Glycobiology, Maureen E. Taylor, Kurt Drickamer, Oxford Univ. Press (2003); Worthington Enzyme Manual, Worthington Biochemical Corp. Freehold, NJ; Handbook of Biochemistry: Section A Proteins Vol I 1976 CRC Press; Handbook of Biochemistry: Section A Proteins Vol II 1976 CRC Press; Essentials of Glycobiology, Cold Spring Harbor Laboratory Press (1999). The nomenclatures used in connection with, and the laboratory procedures and techniques of, biochemistry and molecular biology described herein are those well known and commonly used in the art.

[0043] All publications, patents and other references mentioned herein are incorporated by reference.

[0044] The following terms, unless otherwise indicated, shall be understood to have the following meanings:

[0045] As used herein, the term "N-glycan" refers to an N-linked oligosaccharide, e.g., one that is attached by an asparagine-N-acetylglucosamine linkage to an asparagine residue of a polypeptide. N-glycans have a common pentasaccharide core of $Man_3GlcNAc_2$ ("Man" refers to mannose; "Glc" refers to

glucose; and "NAc" refers to N-acetyl; GlcNAc refers to N-acetylglucosamine). N-glycans differ with respect to the number of branches (antennae) comprising peripheral sugars (e.g., fucose and sialic acid) that are added to the $Man_3GlcNAc_2$ ("Man3") core structure. N-glycans are classified according to their branched constituents (e.g., high mannose, complex or hybrid). A "high mannose" type N-glycan has five or more mannose residues. A "complex" type N-glycan typically has at least one GlcNAc attached to the 1,3 mannose arm and at least one GlcNAc attached to the 1,6 mannose arm of a "trimannose" core. The "trimannose core" is the pentasaccharide core having a Man3 structure. Complex N-glycans may also have galactose ("Gal") residues that are optionally modified with sialic acid or derivatives ("NeuAc", where "Neu" refers to neuraminic acid and "Ac" refers to acetyl). Complex N-glycans may also have intrachain substitutions comprising "bisecting" GlcNAc and core fucose ("Fuc"). A "hybrid" N-glycan has at least one GlcNAc on the terminal of the 1,3 mannose arm of the trimannose core and zero or more mannoses on the 1,6 mannose arm of the trimannose core.

[0046] Abbreviations used herein are of common usage in the art, see, e.g., abbreviations of sugars, above. Other common abbreviations include "PNGase", which refers to peptide N-glycosidase F (EC 3.2.2.18); "GlcNAc Tr (I – III)", which refers to one of three N-acetylglucosaminyltransferase enzymes; "NANA" refers to N-acetylneuraminic acid.

[0047] As used herein, the term "secretion pathway" refers to the assembly line of various glycosylation enzymes to which a lipid-linked oligosaccharide precursor and an N-glycan substrate are sequentially exposed, following the molecular flow of a nascent polypeptide chain from the cytoplasm to the endoplasmic reticulum (ER) and the compartments of the Golgi apparatus. Enzymes are said to be localized along this pathway. An enzyme X that acts on a lipid-linked glycan or an N-glycan before enzyme Y is said to be or to act "upstream" to enzyme Y; similarly, enzyme Y is or acts "downstream" from enzyme X.

[0048] As used herein, the term "antibody" refers to a full antibody (consisting of two heavy chains and two light chains) or a fragment thereof. Such fragments include, but are not limited to, those produced by digestion with various proteases, those produced by chemical cleavage and/or chemical dissociation, and those

produced recombinantly, so long as the fragment remains capable of specific binding to an antigen. Among these fragments are Fab, Fab', F(ab')2, and single chain Fv (scFv) fragments. Within the scope of the term "antibody" are also antibodies that have been modified in sequence, but remain capable of specific binding to an antigen. Example of modified antibodies are interspecies chimeric and humanized antibodies; antibody fusions; and heteromeric antibody complexes, such as diabodies (bispecific antibodies), single-chain diabodies, and intrabodies (see, e.g., Marasco (ed.), Intracellular Antibodies: Research and Disease Applications, Springer-Verlag New York, Inc. (1998) (ISBN: 3540641513), the disclosure of which is incorporated herein by reference in its entirety).

[0049] As used herein, the term "mutation" refers to any change in the nucleic acid or amino acid sequence of a gene product, e.g., of a glycosylation-related enzyme.

[0050] The term "polynucleotide" or "nucleic acid molecule" refers to a polymeric form of nucleotides of at least 10 bases in length. The term includes DNA molecules (e.g., cDNA or genomic or synthetic DNA) and RNA molecules (e.g., mRNA or synthetic RNA), as well as analogs of DNA or RNA containing non-natural nucleotide analogs, non-native internucleoside bonds, or both. The nucleic acid can be in any topological conformation. For instance, the nucleic acid can be single-stranded, double-stranded, triple-stranded, quadruplexed, partially double-stranded, branched, hairpinned, circular, or in a padlocked conformation. The term includes single and double stranded forms of DNA.

[0051] Unless otherwise indicated, a "nucleic acid comprising SEQ ID NO:X" refers to a nucleic acid, at least a portion of which has either (i) the sequence of SEQ ID NO:X, or (ii) a sequence complementary to SEQ ID NO:X. The choice between the two is dictated by the context. For instance, if the nucleic acid is used as a probe, the choice between the two is dictated by the requirement that the probe be complementary to the desired target.

[0052] An "isolated" or "substantially pure" nucleic acid or polynucleotide (e.g., an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, e.g., ribosomes, polymerases, and genomic sequences with which

it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a portion of a polynucleotide in which the "isolated polynucleotide" is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, or (4) does not occur in nature. The term "isolated" or "substantially pure" also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems.

[0053] However, "isolated" does not necessarily require that the nucleic acid or polynucleotide so described has itself been physically removed from its native environment. For instance, an endogenous nucleic acid sequence in the genome of an organism is deemed "isolated" herein if a heterologous sequence (i.e., a sequence that is not naturally adjacent to this endogenous nucleic acid sequence) is placed adjacent to the endogenous nucleic acid sequence, such that the expression of this endogenous nucleic acid sequence is altered. By way of example, a non-native promoter sequence can be substituted (e.g., by homologous recombination) for the native promoter of a gene in the genome of a human cell, such that this gene has an altered expression pattern. This gene would now become "isolated" because it is separated from at least some of the sequences that naturally flank it.

[0054] A nucleic acid is also considered "isolated" if it contains any modifications that do not naturally occur to the corresponding nucleic acid in a genome. For instance, an endogenous coding sequence is considered "isolated" if it contains an insertion, deletion or a point mutation introduced artificially, e.g., by human intervention. An "isolated nucleic acid" also includes a nucleic acid integrated into a host cell chromosome at a heterologous site, a nucleic acid construct present as an episome. Moreover, an "isolated nucleic acid" can be substantially free of other cellular material, or substantially free of culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

[0055] As used herein, the phrase "degenerate variant" of a reference nucleic acid sequence encompasses nucleic acid sequences that can be translated,

according to the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence.

[0056] The term "percent sequence identity" or "identical" in the context of nucleic acid sequences refers to the residues in the two sequences which are the same when aligned for maximum correspondence. The length of sequence identity comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number of different algorithms known in the art which can be used to measure nucleotide sequence identity. For instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson, 1990, (herein incorporated by reference). For instance, percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1, herein incorporated by reference.

[0057] The term "substantial homology" or "substantial similarity," when referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95%, 96%, 97%, 98% or 99% of the nucleotide bases, as measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

[0058] Alternatively, substantial homology or similarity exists when a nucleic acid or fragment thereof hybridizes to another nucleic acid, to a strand of another nucleic acid, or to the complementary strand thereof, under stringent hybridization conditions. "Stringent hybridization conditions" and "stringent wash conditions"

in the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species, length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. One having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization.

[0059] In general, "stringent hybridization" is performed at about 25°C below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of conditions. "Stringent washing" is performed at temperatures about 5°C lower than the T_m for the specific DNA hybrid under a particular set of conditions. The T_m is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. See Sambrook et al., *supra*, page 9.51, hereby incorporated by reference. For purposes herein, "high stringency conditions" are defined for solution phase hybridization as aqueous hybridization (i.e., free of formamide) in 6X SSC (where 20X SSC contains 3.0 M NaCl and 0.3 M sodium citrate), 1% SDS at 65oC for 8-12 hours, followed by two washes in 0.2X SSC, 0.1% SDS at 65oC for 20 minutes. It will be appreciated by the skilled worker that hybridization at 65°C will occur at different rates depending on a number of factors including the length and percent identity of the sequences which are hybridizing.

[0060] The nucleic acids (also referred to as polynucleotides) of this invention may include both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. They may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), pendent moieties (e.g., polypeptides), intercalators (e.g., acridine, psoralen, etc.), chelators, alkylators, and modified linkages (e.g., alpha

anomeric nucleic acids, etc.) Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

[0061] The term "mutated" when applied to nucleic acid sequences means that nucleotides in a nucleic acid sequence may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. A nucleic acid sequence may be mutated by any method known in the art including but not limited to mutagenesis techniques such as "error-prone PCR" (a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. See, e.g., Leung, D. W., et al., *Technique*, 1, pp. 11-15 (1989) and Caldwell, R. C. & Joyce G. F., *PCR Methods Applic.*, 2, pp. 28-33 (1992)); and "oligonucleotide-directed mutagenesis" (a process which enables the generation of site-specific mutations in any cloned DNA segment of interest. See, e.g., Reidhaar-Olson, J. F. & Sauer, R. T., et al., *Science*, 241, pp. 53-57 (1988)).

[0062] The term "vector" as used herein is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome (discussed in more detail below). Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., vectors having an origin of replication which functions in the host cell). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and are thereby replicated along with the host genome. Moreover, certain preferred vectors are capable of directing the

expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression vectors").

[0063] "Operatively linked" expression control sequences refers to a linkage in which the expression control sequence is contiguous with the gene of interest to control the gene of interest, as well as expression control sequences that act in *trans* or at a distance to control the gene of interest.

[0064] The term "expression control sequence" as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination, promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (e.g., ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and transcription termination sequence. The term "control sequences" is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

[0065] The term "recombinant host cell" (or simply "host cell"), as used herein, is intended to refer to a cell into which a recombinant vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein. A recombinant host cell may be an isolated cell or cell line grown in culture or may be a cell which resides in a living tissue or organism.

[0066] The term "peptide" as used herein refers to a short polypeptide, e.g., one that is typically less than about 50 amino acids long and more typically less than about 30 amino acids long. The term as used herein encompasses analogs and mimetics that mimic structural and thus biological function.

[0067] The term "polypeptide" encompasses both naturally-occurring and non-naturally-occurring proteins, and fragments, mutants, derivatives and analogs thereof. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different domains each of which has one or more distinct activities.

[0068] The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide that by virtue of its origin or source of derivation (1) is not associated with naturally associated components that accompany it in its native state, (2) when it exists in a purity not found in nature, where purity can be adjudged with respect to the presence of other cellular material (e.g., is free of other proteins from the same species) (3) is expressed by a cell from a different species, or (4) does not occur in nature (e.g., it is a fragment of a polypeptide found in nature or it includes amino acid analogs or derivatives not found in nature or linkages other than standard peptide bonds). Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be "isolated" from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well known in the art. As thus defined, "isolated" does not necessarily require that the protein, polypeptide, peptide or oligopeptide so described has been physically removed from its native environment.

[0069] The term "polypeptide fragment" as used herein refers to a polypeptide that has an amino-terminal and/or carboxy-terminal deletion compared to a full-length polypeptide. In a preferred embodiment, the polypeptide fragment is a contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally-occurring sequence. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, more

preferably at least 25, 30, 35, 40 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

[0070] A "modified derivative" refers to polypeptides or fragments thereof that are substantially homologous in primary structural sequence but which include, e.g., *in vivo* or *in vitro* chemical and biochemical modifications or which incorporate amino acids that are not found in the native polypeptide. Such modifications include, for example, acetylation, carboxylation, phosphorylation, glycosylation, ubiquitination, labeling, e.g., with radionuclides, and various enzymatic modifications, as will be readily appreciated by those well skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well known in the art, and include radioactive isotopes such as ^{125}I, ^{32}P, ^{35}S, and ^{3}H, ligands which bind to labeled antiligands (e.g., antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods for labeling polypeptides are well known in the art. See Ausubel et al., 1992, hereby incorporated by reference.

[0071] The term "fusion protein" refers to a polypeptide comprising a polypeptide or fragment coupled to heterologous amino acid sequences. Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusion proteins can be produced recombinantly by constructing a nucleic acid sequence which encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the polypeptide or a fragment thereof to another protein.

[0072] The term "non-peptide analog" refers to a compound with properties that are analogous to those of a reference polypeptide. A non-peptide compound may

also be termed a "peptide mimetic" or a "peptidomimetic". See, e.g., Jones, (1992) Amino Acid and Peptide Synthesis, Oxford University Press; Jung, (1997) Combinatorial Peptide and Nonpeptide Libraries: A Handbook John Wiley; Bodanszky et al., (1993) Peptide Chemistry--A Practical Textbook, Springer Verlag; "Synthetic Peptides: A Users Guide", G. A. Grant, Ed, W. H. Freeman and Co., 1992; Evans et al. *J. Med. Chem.* 30:1229 (1987); Fauchere, *J. Adv. Drug Res.* 15:29 (1986); Veber and Freidinger *TINS* p.392 (1985); and references sited in each of the above, which are incorporated herein by reference. Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to useful peptides of the invention may be used to produce an equivalent effect and are therefore envisioned to be part of the invention.

[0073] A "polypeptide mutant" or "mutein" refers to a polypeptide whose sequence contains an insertion, duplication, deletion, rearrangement or substitution of one or more amino acids compared to the amino acid sequence of a native or wild type protein. A mutein may have one or more amino acid point substitutions, in which a single amino acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally-occurring protein, and/or truncations of the amino acid sequence at either or both the amino or carboxy termini. A mutein may have the same but preferably has a different biological activity compared to the naturally-occurring protein.

[0074] A mutein has at least 70% overall sequence homology to its wild-type counterpart. Even more preferred are muteins having 80%, 85% or 90% overall sequence homology to the wild-type protein. In an even more preferred embodiment, a mutein exhibits 95% sequence identity, even more preferably 97%, even more preferably 98% and even more preferably 99%, 99.5% or 99.9% overall sequence identity. Sequence homology may be measured by any common sequence analysis algorithm, such as Gap or Bestfit.

[0075] Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic

activity, and (5) confer or modify other physicochemical or functional properties of such analogs.

[0076] As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. *See Immunology - A Synthesis* (2nd Edition, E.S. Golub and D.R. Gren, Eds., Sinauer Associates, Sunderland, Mass. (1991)), which is incorporated herein by reference. Stereoisomers (e.g., D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as α-, α-disubstituted amino acids, N-alkyl amino acids, and other unconventional amino acids may also be suitable components for polypeptides of the present invention. Examples of unconventional amino acids include: 4-hydroxyproline, γ-carboxyglutamate, ϵ-N,N,N-trimethyllysine, ϵ-N-acetyllysine, O-phosphoserine, N-acetylserine, N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, s-N-methylarginine, and other similar amino acids and imino acids (e.g., 4-hydroxyproline). In the polypeptide notation used herein, the left-hand direction is the amino terminal direction and the right hand direction is the carboxy-terminal direction, in accordance with standard usage and convention.

[0077] A protein has "homology" or is "homologous" to a second protein if the nucleic acid sequence that encodes the protein has a similar sequence to the nucleic acid sequence that encodes the second protein. Alternatively, a protein has homology to a second protein if the two proteins have "similar" amino acid sequences. (Thus, the term "homologous proteins" is defined to mean that the two proteins have similar amino acid sequences). In a preferred embodiment, a homologous protein is one that exhibits 60% sequence homology to the wild type protein, more preferred is 70% sequence homology. Even more preferred are homologous proteins that exhibit 80%, 85% or 90% sequence homology to the wild type protein. In a yet more preferred embodiment, a homologous protein exhibits 95%, 97%, 98% or 99% sequence identity. As used herein, homology between two regions of amino acid sequence (especially with respect to predicted structural similarities) is interpreted as implying similarity in function.

[0078] When "homologous" is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions. A "conservative amino acid substitution" is one in which

an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (e.g., charge or hydrophobicity). In general, a conservative amino acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of homology may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art (see, e.g., Pearson et al., 1994, herein incorporated by reference).

[0079] The following six groups each contain amino acids that are conservative substitutions for one another: 1) Serine (S), Threonine (T); 2) Aspartic Acid (D), Glutamic Acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

[0080] Sequence homology for polypeptides, which is also referred to as percent sequence identity, is typically measured using sequence analysis software. See, e.g., the Sequence Analysis Software Package of the Genetics Computer Group (GCG), University of Wisconsin Biotechnology Center, 910 University Avenue, Madison, Wisconsin 53705. Protein analysis software matches similar sequences using measure of homology assigned to various substitutions, deletions and other modifications, including conservative amino acid substitutions. For instance, GCG contains programs such as "Gap" and "Bestfit" which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild type protein and a mutein thereof. See, e.g., GCG Version 6.1.

[0081] A preferred algorithm when comparing a inhibitory molecule sequence to a database containing a large number of sequences from different organisms is the computer program BLAST (Altschul, S.F. et al. (1990) *J. Mol. Biol.* 215:403-410; Gish and States (1993) *Nature Genet.* 3:266-272; Madden, T.L. et al. (1996) *Meth. Enzymol.* 266:131-141; Altschul, S.F. et al. (1997) *Nucleic Acids Res.*25:3389-3402; Zhang, J. and Madden, T.L. (1997) *Genome Res.* 7:649-656), especially

blastp or tblastn (Altschul et al., 1997). Preferred parameters for BLASTp are: Expectation value: 10 (default); Filter: seg (default); Cost to open a gap: 11 (default); Cost to extend a gap: 1 (default); Max. alignments: 100 (default); Word size: 11 (default); No. of descriptions: 100 (default); Penalty Matrix: BLOWSUM62.

[0082] The length of polypeptide sequences compared for homology will generally be at least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number of different organisms, it is preferable to compare amino acid sequences. Database searching using amino acid sequences can be measured by algorithms other than blastp known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson, 1990, herein incorporated by reference). For example, percent sequence identity between amino acid sequences can be determined using FASTA with its default parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1, herein incorporated by reference.

[0083] The term "domain" as used herein refers to a structure of a biomolecule that contributes to a known or suspected function of the biomolecule. Domains may be co-extensive with regions or portions thereof; domains may also include distinct, non-contiguous regions of a biomolecule. Examples of protein domains include, but are not limited to, an Ig domain, an extracellular domain, a transmembrane domain, and a cytoplasmic domain.

[0084] As used herein, the term "molecule" means any compound, including, but not limited to, a small molecule, peptide, protein, sugar, nucleotide, nucleic acid, lipid, etc., and such a compound can be natural or synthetic.

[0085] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. Exemplary methods and materials are described below, although methods and materials similar or equivalent to those described

herein can also be used in the practice of the present invention and will be apparent to those of skill in the art. All publications and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control. The materials, methods, and examples are illustrative only and not intended to be limiting.

[0086] Throughout this specification and its embodiments, the word "comprise" or variations such as "comprises" or "comprising", will be understood to refer to the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

Nucleic Acid Sequences Encoding Human Endomannosidase Gene

[0087] The rat endomannosidase has been cloned (Spiro et al., *J. Biol. Chem.* 272(46):29356-29363 (1997)). Although the rat endomannosidase is the only cloned member of this family to date, genes and ESTs that show significant homology to this ORF, and in particular to the rat endomannosidase catalytic domain, are in databases. By performing a protein BLAST search using the rat endomannosidase protein sequence (Genbank gi:2642187) we identified two hypothetical human proteins in Genbank having regions of significant homology with the rat endomannosidase sequence (**Example 2**; **Figs. 3A-C**). Combining 5' and 3' regions of these two hypothetical proteins into one ORF produced a putative sequence of 462 amino acids (**Fig. 4**) and a predicted molecular mass of 54 kDa. Alignment of this putative human endomannosidase sequence to the known rat sequence indicated that the C-termini of these proteins are highly conserved but that the N-termini are more varied (**Fig. 7**). It is likely that the conserved region (i.e., from the motif 'DFQ(K/R)SDRIN' to the C-terminus), corresponds to the catalytic domain in each endomannosidase, or at least to a region essential for activity.

[0088] Based on the above-deduced human endomannosidase gene sequence, we constructed primers and amplified an open reading frame (ORF) from a human liver cDNA library by PCR (**Example 2**). The nucleic acid sequence which encodes that ORF is 77.8% identical across its length to the full-length nucleic acid sequence encoding the rat endomannosidase ORF (sequence pair distances using

the Clustal methods with weighted residue weight table). At the amino acid sequence level, the human and rat endomannosidase proteins are predicted to be 76.7% identical overall. In the more conserved region noted above (i.e., from the motif 'DFQ(K/R)SDRIN' to the C-terminus), the proteins are 86.6% identical overall. Unlike the rat protein, the predicted human protein has a very hydrophobic region at the N-terminus (residues 10 to 26) which may be a transmembrane region (**Fig. 4**, boxed). The human endomannosidase (unlike the rat protein), is predicted to be a type-II membrane protein, as are most other higher eukaryotic mannosidases.

[0089] We subcloned the human endomannosidase ORF into various vectors, including a yeast integration plasmid (**Example 3**), to study the effect of its expression on the N-glycosylation pathway of a lower eukaryotic host cell, *Pichia pastoris*. As described below, engineering the human mannosidase enzyme into the glycosylation pathway of a host cell significantly affects the subsequent glycosylation profile of proteins produced in that host cell and its descendants. Preferably, the host cell is engineered to express a human mannosidase enzyme activity (e.g., from a catalytic domain) in combination with one or more other engineered glycosylation activities to make human-like glycoproteins.

[0090] Accordingly, the present invention provides isolated nucleic acid molecules, including but not limited to nucleic acid molecules comprising or consisting of a full-length nucleic acid sequence encoding human endomannosidase. The nucleic acid sequence and the ORF of human endomannosidase are set forth in **Fig. 4** and as SEQ ID NO:1. The encoded amino acid sequence is also set forth in **Fig. 4** and in SEQ ID NO:2.

[0091] In one embodiment, the invention provides isolated nucleic acid molecules having a nucleic acid sequence comprising or consisting of a wild-type human endomannosidase coding sequence (SEQ ID NO:1); homologs, variants and derivatives thereof; and fragments of any of the above. In one embodiment, the invention provides a nucleic acid molecule comprising or consisting of a sequence which is a degenerate variant of the wild-type human endomannosidase coding sequence (SEQ ID NO:1). In a preferred embodiment, the invention provides a nucleic acid molecule comprising or consisting of a sequence which is a variant of

the human endomannosidase coding sequence (SEQ ID NO:1) having at least 65% identity to the wild-type gene. The nucleic acid sequence can preferably have at least 70%, 75% or 80% identity to the wild-type human endomannosidase coding sequence (SEQ ID NO:1) (specifically excluding, however, the rat endomannosidase gene, which is about 78% identical overall). Even more preferably, the nucleic acid sequence can have 85%, 90%, 95%, 98%, 99%, 99.9%, or higher, identity to the wild-type human endomannosidase coding sequence (SEQ ID NO:1).

[0092] In another embodiment, the nucleic acid molecule of the invention encodes a polypeptide comprising or consisting of the amino acid sequence of SEQ ID NO:2. Also provided is a nucleic acid molecule encoding a polypeptide sequence that is at least 65% identical to SEQ ID NO:2 (specifically excluding, however, the rat endomannosidase polypeptide, which is about 77% identical overall). Typically the nucleic acid molecule of the invention encodes a polypeptide sequence of at least 70%, 75% or 80% identity to SEQ ID NO:2. Preferably, the encoded polypeptide is at least 85%, 90% or 95% identical to SEQ ID NO:2, and the identity can even more preferably be 98%, 99%, 99.9% or even higher.

[0093] The invention also provides nucleic acid molecules that hybridize under stringent conditions to the above-described nucleic acid molecules. As defined above, and as is well known in the art, stringent hybridizations are performed at about 25 °C below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of conditions, where the T_m is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. Stringent washing is performed at temperatures about 5 °C lower than the T_m for the specific DNA hybrid under a particular set of conditions.

[0094] Nucleic acid molecules comprising a fragment of any one of the above-described nucleic acid sequences are also provided. These fragments preferably contain at least 20 contiguous nucleotides. More preferably the fragments of the nucleic acid sequences contain at least 25, 30, 35, 40, 45 or 50 contiguous nucleotides. Even more preferably, the fragments of the nucleic acid sequences contain at least 60, 70, 80, 90, 100 or more contiguous nucleotides. In a further

embodiment of the invention, the nucleic acid sequence is a variant of the fragment having at least 65% identity to the wild-type gene fragment. The nucleic acid sequence can preferably have at least 70%, 75% or 80% identity to the wild-type gene fragment. Even more preferably, the nucleic acid sequence can have 85%, 90%, 95%, 98%, 99%, 99.9% or even higher identity to the wild-type gene fragment.

[0095] The nucleic acid sequence fragments of the present invention display utility in a variety of systems and methods. For example, the fragments may be used as probes in various hybridization techniques. Depending on the method, the target nucleic acid sequences may be either DNA or RNA. The target nucleic acid sequences may be fractionated (*e.g.*, by gel electrophoresis) prior to the hybridization, or the hybridization may be performed on samples *in situ*. One of skill in the art will appreciate that nucleic acid probes of known sequence find utility in determining chromosomal structure (*e.g.*, by Southern blotting) and in measuring gene expression (*e.g.*, by Northern blotting). In such experiments, the sequence fragments are preferably detectably labeled, so that their specific hydridization to target sequences can be detected and optionally quantified. One of skill in the art will appreciate that the nucleic acid fragments of the present invention may be used in a wide variety of blotting techniques not specifically described herein.

[0096] It should also be appreciated that the nucleic acid sequence fragments disclosed herein also find utility as probes when immobilized on microarrays. Methods for creating microarrays by deposition and fixation of nucleic acids onto support substrates are well known in the art. Reviewed in *DNA Microarrays : A Practical Approach (Practical Approach Series)*, Schena (ed.), Oxford University Press (1999) (ISBN: 0199637768); *Nature Genet.* 21(1)(suppl):1-60 (1999); *Microarray Biochip: Tools and Technology*, Schena (ed.), Eaton Publishing Company/BioTechniques Books Division (2000) (ISBN: 1881299376), the disclosures of which are incorporated herein by reference in their entireties. Analysis of, for example, gene expression using microarrays comprising nucleic acid sequence fragments, such as the nucleic acid sequence fragments disclosed herein, is a well-established utility for sequence fragments in the field of cell and

molecular biology. Other uses for sequence fragments immobilized on microarrays are described in Gerhold *et al., Trends Biochem. Sci.* 24:168-173 (1999) and Zweiger, *Trends Biotechnol.* 17:429-436 (1999); *DNA Microarrays : A Practical Approach (Practical Approach Series)*, Schena (ed.), Oxford University Press (1999) (ISBN: 0199637768); *Nature Genet.* 21(1)(suppl):1-60 (1999); *Microarray Biochip: Tools and Technology*, Schena (ed.), Eaton Publishing Company/BioTechniques Books Division (2000) (ISBN: 1881299376), the disclosures of each of which is incorporated herein by reference in its entirety. In another embodiment, isolated nucleic acid molecules encoding a polypeptide having endomannosidase activity are provided. As is well known in the art, enzyme activities can be measured in various ways. Alternatively, the activity of the enzyme can be followed using chromatographic techniques, such as by high performance liquid chromatography. Chung and Sloan, *J. Chromatogr.* 371:71-81 (1986). Other methods and techniques may also be suitable for the measurement of enzyme activity, as would be known by one of skill in the art.

[0097] In another embodiment, the nucleic acid molecule of the invention encodes a polypeptide having the amino acid sequence of SEQ ID NO:2. The nucleic acid sequence of the invention encodes a polypeptide having at least 77% identity to the wild-type rat endomannosidase gene (Genbank AF023657). In another embodiment, the nucleic acid sequence has at least 87% identity to the wild-type rat endomannosidase catalytic domain. In an even more preferred embodiment, the nucleic acid sequence can have 90%, 95%, 98%, 99%, 99.9% or even higher identity to the wild-type rat endomannosidase gene.

[0098] Polypeptides encoded by the nucleic acids of the invention, especially peptides having a biological (e.g., catalytic or other) and/or immunological activity, are also provided by the invention.

Nucleic Acid Sequences Encoding Mouse Endomannosidase Gene

[0099] The mouse endomannosidase gene is cloned by designing primers that complement the putative homologous regions between the mouse and human endomannosidase genes and PCR amplifying to generate a probe which can be used to pull out a full-length cDNA encoding mouse endomannosidase (**Example**

2). The nucleotide and predicted amino acid sequence of the mouse endomannosidase open reading frame (ORF) is set forth in **Fig. 6** and as SEQ ID NOs:3 and 4, respectively.

[0100] The mouse ORF shows substantial homology to the known rat endomannosidase and the human liver endomannosidase of the present invention (**Fig. 7**). Specifically, the nucleic acid sequence which encodes the mouse endomannosidase ORF is 86.0% and 84.2% identical across its length to the full-length nucleic acid sequence encoding the rat and the human endomannosidase ORFs, respectively (sequence pair distances using the Clustal methods with weighted residue wieight table). At the amino acid sequence level, the mouse and rat endomannosidase proteins are predicted to be 82.3% identical, amd the mouse and human endomannosidase proteins are predicted to be 84.9% identical overall. In the more conserved region noted above (i.e., from the motif 'DFQ(K/R)SDRIN' to the C-terminus), the mouse and rat proteins are 92.3% identical, and the mouse and human proteins are 86.1% identical, overall.

[0101] Accordingly, the present invention further provides isolated nucleic acid molecules and variants thereof encoding the mouse endomannosidase. In one embodiment, the invention provides an isolated nucleic acid molecule having a nucleic acid sequence comprising or consisting of the gene encoding the mouse endomannosidase (SEQ ID NO:3), homologs, variants and derivatives thereof.

[0102] Accordingly, the present invention provides isolated nucleic acid molecules, including but not limited to nucleic acid molecules comprising or consisting of a full-length nucleic acid sequence encoding mouse endomannosidase. The nucleic acid sequence and the ORF of mouse endomannosidase are set forth in **Fig. 6** and as SEQ ID NO:3. The encoded amino acid sequence is also set forth in **Fig. 6** and in SEQ ID NO:4.

[0103] In one embodiment, the invention provides isolated nucleic acid molecules having a nucleic acid sequence comprising or consisting of a wild-type mouse endomannosidase coding sequence (SEQ ID NO:3); homologs, variants and derivatives thereof; and fragments of any of the above. In one embodiment, the invention provides a nucleic acid molecule comprising or consisting of a sequence which is a degenerate variant of the wild-type mouse endomannosidase coding

sequence (SEQ ID NO:3). In a preferred embodiment, the invention provides a nucleic acid molecule comprising or consisting of a sequence which is a variant of the mouse endomannosidase coding sequence (SEQ ID NO:3) having at least 65% identity to the wild-type gene. The nucleic acid sequence can preferably have at least 70%, 75%, 80% or 85% identity to the wild-type human endomannosidase coding sequence (SEQ ID NO:3) (specifically excluding, however, the rat endomannosidase gene, which is about 86% identical overall). Even more preferably, the nucleic acid sequence can have 90%, 95%, 98%, 99%, 99.9%, or higher, identity to the wild-type mouse endomannosidase coding sequence (SEQ ID NO:3).

[0104] In another embodiment, the nucleic acid molecule of the invention encodes a polypeptide comprising or consisting of the amino acid sequence of SEQ ID NO:4. Also provided is a nucleic acid molecule encoding a polypeptide sequence that is at least 65% identical to SEQ ID NO:4 (specifically excluding, however, the rat endomannosidase polypeptide, which is about 82% identical overall). Typically the nucleic acid molecule of the invention encodes a polypeptide sequence of at least 70%, 75% or 80% identity to SEQ ID NO:4. Preferably, the encoded polypeptide is at least 85%, 90% or 95% identical to SEQ ID NO:4, and the identity can even more preferably be 98%, 99%, 99.9% or even higher.

[0105] The invention also provides nucleic acid molecules that hybridize under stringent conditions to the above-described nucleic acid molecules. As defined above, and as is well known in the art, stringent hybridizations are performed at about 25°C below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of conditions, where the T_m is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. Stringent washing is performed at temperatures about 5°C lower than the T_m for the specific DNA hybrid under a particular set of conditions.

[0106] Nucleic acid molecules comprising a fragment of any one of the above-described nucleic acid sequences are also provided. These fragments preferably contain at least 20 contiguous nucleotides. More preferably the fragments of the nucleic acid sequences contain at least 25, 30, 35, 40, 45 or 50 contiguous

nucleotides. Even more preferably, the fragments of the nucleic acid sequences contain at least 60, 70, 80, 90, 100 or more contiguous nucleotides. In a further embodiment of the invention, the nucleic acid sequence is a variant of the fragment having at least 65% identity to the wild-type gene fragment. The nucleic acid sequence can preferably have at least 70%, 75% or 80% identity to the wild-type gene fragment. Even more preferably, the nucleic acid sequence can have 85%, 90%, 95%, 98%, 99%, 99.9% or even higher identity to the wild-type gene fragment.

[0107] In another embodiment, the nucleic acid molecule of the invention encodes a polypeptide comprising or consisting of the amino acid sequence of SEQ ID NO:4. Also provided is a nucleic acid molecule encoding a polypeptide sequence that is at least 65% identical to SEQ ID NO:4 (specifically excluding, however, the rat endomannosidase polypeptide, which is about 82% identical overall). Typically the nucleic acid molecule of the invention encodes a polypeptide sequence of at least 70%, 75% or 80% identity to SEQ ID NO:4. Preferably, the encoded polypeptide is at least 85%, 90% or 95% identical to SEQ ID NO:4, and the identity can even more preferably be 98%, 99%, 99.9% or even higher.

[0108] In a preferred embodiment, the nucleic acid molecule of the invention encodes a polypeptide having at least 83% identity to the wild-type rat endomannosidase gene (Genbank AF023657). In another embodiment, the nucleic acid sequence encoding an amino acid sequence has at least 93% identity to the wild-type rat endomannosidase catalytic domain. In an even more preferred embodiment, the nucleic acid sequence can have 94%, 95%, 98%, 99%, 99.9% or even higher identity to the wild-type rat endomannosidase gene.

[0109] Polypeptides encoded by the nucleic acids of the invention, especially peptides having a biological (e.g., catalytic or other) and/or immunological activity, are also provided by the invention.

Characterization of Encoded Endomannosidase Products

[0110] The human liver endomannosidase and the putative mouse endomannosidase are the second and third members of a newly developing family

of glycosidic enzymes, with the rat endomannosidase enzyme being the first such member. Sequence comparison of the human, mouse and rat ORFs (**Fig. 7**) demonstrates high homology from the motif 'DFQ(K/R)SDRI' to the C-termini of the sequences suggesting that this region encodes an essential fragment of the protein, and potentially, the catalytic domain. In contrast, the lower homology within the N-termini of the proteins demonstrates evolutionary divergence. Like the majority of glycosidases and glycosyltransferases, the mouse and human enzymes have a hydrophobic region indicative of a transmembrane domain. Such a domain would facilitate the orientation and localization of the enzyme in the secretory pathway. In contrast, the rat endomannosidase does not have a transmembrane domain but does have a glycine residue at position 2 (Spiro 1997, *supra*). This penultimate glycine residue has the potential to be myristoylated which in turn provides a mechanism for membrane localization (Boutin, *Cell Signal* 9: 15-35 (1997)). Alternatively, myristoylation may not be the means of rat endomannosidase localization to the Golgi (Zuber 2000, *supra*) -- protein-protein interactions may be the determining mechanism.

[0111] Like the rat endomannosidase, both the human and mouse isoforms are predicted to localize to the Golgi based on the activity of this class of proteins. Traditionally, the removal of glucose from N-glycans was thought to occur in the ER by glucosidases I and II. However, the characterization of endomannosidase and its localization to the *cis* and *medial* cisternae of the Golgi demonstrates that glucose trimming does occur subsequent to glucosidase localization (Roth et al. *Biochimie* 85: 287-294 (2003)).

[0112] The specific role that endomannosidase fulfills is currently uncertain. Affinity-purification of rat endomannosidase demonstrated the co-purification with calreticulin suggesting its role in the quality control of N-glycosylation (Spiro et al., *J. Biol. Chem.* 271: 11588-11594 (1996)). Alternatively, endomannosidase may provide the cell with the ability to recover and properly mature glucosylated structures that have by-passed glucosidase trimming. Thus, removing the glucose-α1,3-mannose dimer from a glucosylated high mannose structure presents a substrate for the resident Golgi glycosidic and glycosyltransferase enzymes, enabling the maturation of the N-glycans.

[0113] We analyzed the tissue distribution of human endomannosidase and, like the rat isoform (Spiro (1997)), it was widespread in the tissues examined (**Fig. 8**) (**Example 6**). The liver and kidney demonstrated high expression levels but the pattern in the remainder of the tissues was significantly different. Interestingly, in contrast to the human endomannosidase, the rat isoform shows high expression levels in both the brain and lung (Spiro (1997)). The widespread expression of both isoforms of this enzyme in rat and human suggests that endomannosidase may play a house-keeping role in the processing of N-glycans.

[0114] Expression in *P. pastoris* of the human endomannosidase of the invention confirms that the isolated ORF has activity. Interestingly, the rat isoform, though highly homologous at the nucleotide and protein levels, is expressed at levels at least five-fold higher than the human protein as seen on Western Blots (**Fig. 9**). It is possible that rat enzyme is inherently more stable during expression or in the culture medium.

[0115] Both recombinantly expressed endomannosidase enzymes were processed at their C-termini. In the case of the human enzyme, C-terminal processing appeared to be complete (based on apparent total conversion of the 59kDa band to the 54kDa form, presumably due to the lower expression level). In contrast, though the majority of the rat isoform was the 54kDa form, some of the 59kDa band remained (**Example 7**). Likewise, when the rat endomannosidase was expressed in *Escherichia coli,* the protein was proteolytically processed at the C-terminus over time (Spiro 1997, *supra*). Furthermore, affinity chromatographic purification of the rat isoform from rat liver demonstrated the presence of two forms, 56 and 60 kDa (Hiraizumi et al., *J. Biol. Chem.* 269: 4697-4700 (1994)). Together, these data indicate that both the human and rat endomannosidase proteins are susceptible to proteolytic processing. Based on the similar sizes of the two enzymes following proteolysis, the cleavage site is likely the same. Whether the cleavage site in the bacterial, yeast and mammalian systems is the same remains to be determined. Further characterization of the endomannosidase shows an optimal activity at about pH 6.2 (**Example 9**) and a temperature optimum of about 37°C (**Example 9**).

[0116] The isolation and characterization of the human endomannosidase and the identification of the mouse homologue expands this family of glycosidases from a solitary member consisting of the rat isoform. This in turn has allowed us to characterize further this family of proteins. Indeed, this has allowed us to demonstrate that, while the C-terminal sequences of these proteins are highly conserved, variations in the N-terminal architecture occur. A previously reported phylogenetic survey of endomannosidase indicated that this protein has emerged only recently during evolution and is restricted to members of the chordate phylum, which includes mammals, birds, reptiles, amphibians and bony fish, with the only exception being that it has also been identified in Mollusca (Dairaku and Spiro, *Glycobiology* 7: 579-586 (1997)). Therefore, the isolation of more diversified members of this family of proteins will expectedly demonstrate further variations in endomannosidase structure and, potentially, activity.

Utility of Endomannosidase Expression

[0117] The human and mouse endomannosidase enzymes or catalytic domains (and nucleic acid molecules of the invention encoding such activities) will each be useful, e.g., for modifying certain glycosylation structures, in particular, for hydrolyzing a composition comprising at least one glucose residue and one mannose residue on a glucosylated glycan structure (**Fig. 1** and **Fig. 2**). In one embodiment, the encoded enzyme catalyzes the cleavage of a di- tri-, or tetra-saccharide composition comprising at least one glucose residue and one mannose residue of glucosylated glycan precursors (**Fig. 1**). In another embodiment, the encoded enzyme also modifies a number of glucosylated structures, including $Glc_{1-3}Man_{9-5}GlcNAc_2$ (**Fig. 2**). One or more nucleic acids and/or polypeptides of the invention are introduced into a host cell of choice to modify the glycoproteins produced by that host cell.

Cellular Targeting of Endomannosidase *In Vivo*

[0118] Although glucosidases act upon high mannan glycans in the ER, some mannans escape the ER without proper modification and, thus, mannans with undesired glycosylations move through the secretory pathway. Previous studies

suggest that in higher eukaryotes a fraction of glucosylated mannose structures does bypass the quality control of the ER, and that endomannosidase is present in the subsequent compartment to recover this fraction. Accordingly, in a feature of the present invention, the endomannosidase modifies the glucosylated mannose structures that have bypassed the ER. In a preferred embodiment, the endomannosidase enzyme encoded by the nucleic acid of the present invention is localized in the Golgi, trans Golgi network, transport vesicles or the ER. The enzymes are involved in the trimming of glucosylated high mannan glycans in yeast. For example, the glucosylated structure $GlcMan_9GlcNAc_2$, which has bypassed the ER glucosidase I and II enzymes, is modified by the endomannosidase in which at least a glucose-mannose residue is hydrolyzed producing $Man_8GlcNAc_2$. The endomannosidase enzymes of the present invention act as a quality control step in the Golgi, recovering the glucosylated high mannan glycans and removing a composition comprising at least one glucose residue and one mannose residue.

Combinatorial Nucleic Acid Library Encoding
Endomannosidase Catalytic Domains

[0119] In another aspect of the invention, one or more chimeric nucleic acid molecules encoding novel endomannosidase proteins is constructed by forming a fusion protein between an endomannosidase enzyme and a cellular targeting signal peptide, e.g., by the in-frame ligation of a DNA fragment encoding a cellular targeting signal peptide with a DNA fragment encoding an endomannosidase enzyme or catalytically active fragment thereof. Preferably, one or more fusion proteins are made in the context of an endomannosidase combinatorial DNA library. See generally WO 02/00879 and the publication of United States Application No. 10/371,877 (filed Feb. 20, 2003); each of which is incorporated herein by reference in nits entirety. The endomannosidase DNA library comprises a wide variety of fusion constructs, which are expressed in a host cell of interest, e.g., by using an integration plasmid such as the pRCD259 (**Example 5**).

Targeting Peptide Sequence Sub-Libraries

[0120] Another useful sub-library includes nucleic acid sequences encoding targeting signal peptides that result in localization of a protein to a particular location within the ER, Golgi, or trans Golgi network. These targeting peptides may be selected from the host organism to be engineered as well as from other related or unrelated organisms. Generally such sequences fall into three categories: (1) N-terminal sequences encoding a cytosolic tail (ct), a transmembrane domain (tmd) and part or all of a stem region (sr), which together or individually anchor proteins to the inner (lumenal) membrane of the Golgi; (2) retrieval signals which are generally found at the C-terminus such as the HDEL or KDEL tetrapeptide; and (3) membrane spanning regions from various proteins, e.g., nucleotide sugar transporters, which are known to localize in the Golgi.

[0121] In the first case, where the targeting peptide consists of various elements (cytosolic tail (ct), transmembrane domain (tmd) and stem region (sr)), the library is designed such that the ct, the tmd and various parts of the stem region are represented. Accordingly, a preferred embodiment of the sub-library of targeting peptide sequences includes ct, tmd, and/or sr sequences from membrane-bound proteins of the ER or Golgi. In some cases it may be desirable to provide the sub-library with varying lengths of sr sequence. This may be accomplished by PCR using primers that bind to the 5' end of the DNA encoding the cytosolic region and employing a series of opposing primers that bind to various parts of the stem region.

[0122] Still other useful sources of targeting peptide sequences include retrieval signal peptides, e.g. the tetrapeptides HDEL or KDEL, which are typically found at the C-terminus of proteins that are transported retrograde into the ER or Golgi. Still other sources of targeting peptide sequences include (a) type II membrane proteins, (b) the enzymes with optimum pH, (c) membrane spanning nucleotide sugar transporters that are localized in the Golgi, and (d) sequences referenced in **Table 1**.

Table 1. Sources of useful compartmental targeting sequences

Gene or Sequence	**Organism**	**Function**	**Location of Gene Product**
MNS1	*A.nidulans*	α-1,2-mannosidase	ER
MNS1	*A.niger*	α-1,2-mannosidase	ER
MNS1	*S.cerevisiae*	α-1,2-mannosidase	ER
GLS1	*S.cerevisiae*	glucosidase	ER
GLS1	*A.niger*	glucosidase	ER
GLS1	*A.nidulans*	glucosidase	ER
HDEL at C-terminus	*Universal in fungi*	retrieval signal	ER
SEC12	*S.cerevisiae*	COPII vesicle protein	ER/Golgi
SEC12	*A.niger*	COPII vesicle protein	ER/Golgi
OCH1	*S.cerevisiae*	1,6-mannosyltransferase	Golgi (cis)
OCH1	*P.pastoris*	1,6-mannosyltransferase	Golgi (cis)
MNN9	*S.cerevisiae*	1,6-mannosyltransferase complex	Golgi
MNN9	*A.niger*	undetermined	Golgi
VAN1	*S.cerevisiae*	undetermined	Golgi
VAN1	*A.niger*	undetermined	Golgi
ANP1	*S.cerevisiae*	undetermined	Golgi
HOC1	*S.cerevisiae*	undetermined	Golgi
MNN10	*S.cerevisiae*	undetermined	Golgi
MNN10	*A.niger*	undetermined	Golgi
MNN11	*S.cerevisiae*	undetermined	Golgi (cis)
MNN11	*A.niger*	undetermined	Golgi (cis)
MNT1	*S.cerevisiae*	1,2-mannosyltransferase	Golgi (cis, medial
KTR1	*P.pastoris*	undetermined	Golgi (medial)
KRE2	*P.pastoris*	undetermined	Golgi (medial)
KTR3	*P.pastoris*	undetermined	Golgi (medial)
MNN2	*S.cerevisiae*	1,2-mannosyltransferase	Golgi (medial)

Gene or Sequence	**Organism**	**Function**	**Location of Gene Product**
KTR1	*S.cerevisiae*	undetermined	Golgi (medial)
KTR2	*S.cerevisiae*	undetermined	Golgi (medial)
MNN1	*S.cerevisiae*	1,3-mannosyltransferase	Golgi (trans)
MNN6	*S.cerevisiae*	Phosphomannosyltransferase	Golgi (trans)
2,6 ST	*H. sapiens*	2,6-sialyltransferase	trans Golgi network
UDP-Gal T	*S. pombe*	UDP-Gal transporter	Golgi

Endomannosidase Fusion Constructs

[0123] A representative example of an endomannosidase fusion construct derived from a combinatorial DNA library of the invention inserted into a plasmid is **pSH280**, which comprises a truncated *Saccharomyces MNN11*(m) targeting peptide (1-303 nucleotides of *MNN11* from SwissProt P46985), constructed from primers SEQ ID NO: 5 and SEQ ID NO: 6, ligated in-frame to a 48 N-terminal amino acid deletion of a rat endo-α1,2-mannosidase (Genbank AF 023657). The nomenclature used herein, thus, refers to the targeting peptide/catalytic domain region of a glycosylation enzyme as *Saccharomyces MNN11*(m)/rat endomannosidase Δ48. The encoded fusion protein localizes in the Golgi by means of the *MNN11* targeting peptide sequence while retaining its endomannosidase catalytic domain activity and is capable of producing unglucosylated N-glycans such as $Man_4GlcNAc_2$ in a lower eukaryote. The glycan profile from a reporter glycoprotein K3 expressed in a strain of *P. pastoris* **RDP25** (*och1 alg3*) transformed with **pSH280** exhibits a peak, among others, at 1099 m/z **[c]** corresponding to the mass of $Man_4GlcNAc_2$ and 1424 m/z **[a]** corresponding to the mass of hexose 6 (**Fig. 10B**; see Examples 11 and 12). This new *P. pastoris* strain, designated as **YSH97,** shows greater than about 95% endomannosidase activity evidenced by the extent to which the glucosylated hexose 6 structure is removed from the reporter glycoprotein.

[0124] The structure of hexose 6 **[a]** expressed in a host cell (e.g., *P. pastoris* **RDP25**) comprises a mixture of glycans comprising $GlcMan_5GlcNAc_2$ and $Man_6GlcNAc_2$ and its isomers (**Fig. 10A**). By introduction and expression of the

endomannosidase of the present invention in a host cell, a composition comprising at least one glucose residue and mannose residue is removed from the hexose 6 structure (**Fig. 10B**). The glucosylated structure $GlcMan_5GlcNAc_2$ is readily converted to $Man_4GlcNAc_2$, which is then subsequently converted to $Man_3GlcNAc_2$ with α1,2-mannosidase *in vitro* digestion. The hexose 6 species comprising the glucosylated mannans is not cleaved by α1,2-mannosidase. The predominant peak corresponding to the structure $Man_3GlcNAc_2$ **[b]** (**Fig. 10C**) shown after the α1,2-mannosidase digestion confirms the apparent removal of the glucose-mannose dimer from $GlcMan_5GlcNAc_2$ exposing a terminal Manα1,2 on $Man_4GlcNAc_2$ for hydrolysis producing $Man_3GlcNAc_2$.

[0125] The other species of hexose 6: $Man_6GlcNAc_2$ is not readily affected by the endomannosidase of the present invention and accordingly, is contemplated as un-glucosylated structures. A skilled artisan would appreciate that this species of hexose 6: $Man_6GlcNAc_2$ comprises Manα1,2 additions, which is evidenced by the subsequent α1,2-mannosidase *in vitro* digestion producing $Man_3GlcNAc_2$ (**Fig. 10C**).

[0126] Another example of an endomannosidase fusion construct derived from a combinatorial DNA library of the invention inserted into a plasmid is **pSH279**, which is a truncated *Saccharomyces VAN1*(s) targeting peptide (1-279 nucleotides of *VAN1* from SwissProt P23642) constructed from primers SEQ ID NO: 7 and SEQ ID NO: 8, ligated in-frame to a 48 N-terminal amino acid deletion of a rat endo-α1,2-mannosidase (Genbank AF 023657). The nomenclature used herein, thus, refers to the targeting peptide/catalytic domain region of a glycosylation enzyme as *Saccharomyces VAN1*(s)/rat endomannosidase Δ48. The encoded fusion protein localizes in the Golgi by means of the *VAN1* targeting peptide sequence while retaining its endomannosidase catalytic domain activity and is capable of producing N-glycans having a $Man_4GlcNAc_2$ structure in *P. pastoris* (**RDP25**). The glycan profile from a reporter glycoprotein K3 expressed in a strain of *P.pastoris* **RDP-25** (*och1 alg3*) transformed with **pSH279** exhibits a peak, among others, at 1116 m/z **[c]** corresponding to the mass of $Man_4GlcNAc_2$ and 1441 m/z **[a]** corresponding to the mass of hexose 6 (**Fig. 11**; examples 11 and 12). **Fig. 11B** shows a residual hexose 6 **[a]** peak indicating only partial activity of the

endomannosidase. This strain, designated as **YSH96**, shows greater than about 40% endomannosidase activity, evidenced by the extent to which the glucosylated hexose 6 structure is removed from the reporter glycoprotein.

[0127] The structure of hexose 6 **[a]** expressed in a host cell (e.g., *P. pastoris* **RDP25**) comprises a mixture of glycans comprising $GlcMan_5GlcNAc_2$ and $Man_6GlcNAc_2$ and its isomers (**Fig. 11A**). By introduction and expression of the endomannosidase of the present invention in a host cell, a composition comprising at least one glucose residue and mannose residue is removed from the hexose 6 structure (**Fig. 11B**). The glucosylated structure $GlcMan_5GlcNAc_2$ is readily converted to $Man_4GlcNAc_2$, which is then subsequently converted to $Man_3GlcNAc_2$ with α1,2-mannosidase *in vitro* digestion. The hexose 6 species comprising the glucosylated mannans is not cleaved by α1,2-mannosidase. The predominant peak corresponding to the structure $Man_3GlcNAc_2$ **[b]** (**Fig. 11C**) shown after the α1,2-mannosidase digestion confirms the apparent removal of the glucose-mannose dimer from $GlcMan_5GlcNAc_2$ exposing a terminal Manα1,2 on $Man_4GlcNAc_2$ for hydrolysis producing $Man_3GlcNAc_2$.

[0128] The other species of hexose 6: $Man_6GlcNAc_2$ is not readily affected by the endomannosidase of the present invention and accordingly, is contemplated as un-glucosylated structures. A skilled artisan would appreciate that this species of hexose 6: $Man_6GlcNAc_2$ comprises Manα1,2 additions, which is evidenced by the subsequent α1,2-mannosidase *in vitro* digestion producing $Man_3GlcNAc_2$ (**Fig. 11C**).

[0129] Additionally, an example of an endomannosidase fusion construct inserted into a plasmid that does not show apparent catalytic activity derived from a combinatorial DNA library of the invention is **pSH278**, which a truncated *Saccharomyces GLS1*(s) targeting peptide (1-102 nucleotides of *GLS1* from SwissProt P53008) constructed from primers SEQ ID NO: 9 and SEQ ID NO: 10, ligated in-frame to a 48 N-terminal amino acid deletion of a rat endo-α1,2-mannosidase (Genbank AF 023657). The nomenclature used herein, thus, refers to the targeting peptide/catalytic domain region of a glycosylation enzyme as *Saccharomyces GLS1*(s)/rat endomannosidase Δ48. The glycan profile from a reporter glycoprotein K3 expressed in a strain of a *P.pastoris* **RDP-25** (*och1 alg3*)

transformed with **pSH278** exhibits, a peak, among others, at 1439 m/z (K^+ adduct) **[c]** and a peak at 1422 m/z (Na^+ adduct) corresponding to the mass of hexose 6 **[a]** (**Fig. 12;** examples 11 and 12). This strain, designated as **YSH95,** shows less than about 10% endomannosidase activity as evidenced by the extent to which the glucosylated hexose 6 structure is removed from the reporter glycoprotein.

[0130] Unlike the previous two glycan profiles shown in **Figs. 10 and 11**, the endomannosidase construct **pSH278** expressed in *P. pastoris* **RDP25** shows relatively low endomannosidase activity (**Fig. 12**). Subsequent digestion with α1,2 mannosidase, however, reveals a peak corresponding to the mass of $Man_3GlcNAc_2$ **[b]**. A skilled artisan would appreciate that the hexose 6 species comprising $Man_6GlcNAc_2$ have been converted to $Man_3GlcNAc_2$ by introduction of α1,2 mannosidase whereas the other hexose 6 species comprising $GlcMan_5GlcNAc_2$ are still present, which, in effect, are still glucosylated.

[0131] By creating a combinatorial DNA library of these and other such endomannosidase fusion constructs according to the invention, a skilled artisan may distinguish and select those constructs having optimal intracellular endomannosidase trimming activity from those having relatively low or no activity. Methods using combinatorial DNA libraries of the invention are advantageous because only a select few endomannosidase fusion constructs may produce a particularly desired *N*-glycan *in vivo*. In addition, endomannosidase trimming activity may be specific to a particular protein of interest. Thus, it is to be further understood that not all targeting peptide/mannosidase catalytic domain fusion constructs may function equally well to produce the proper glycosylation on a glycoprotein of interest. Accordingly, a protein of interest may be introduced into a host cell transformed with a combinatorial DNA library to identify one or more fusion constructs which express a mannosidase activity optimal for the protein of interest. One skilled in the art will be able to produce and select optimal fusion construct(s) using the combinatorial DNA library approach described herein.

[0132] It is apparent, moreover, that other such fusion constructs exhibiting localized active endomannosidase catalytic domains may be made using techniques such as those exemplified in WO 02/00879 and described herein. It will be a

matter of routine experimentation for one skilled in the art to make and use the combinatorial DNA library of the present invention to optimize non-glucosylated N-glycans (for example $Man_4GlcNAc_2$) production from a library of fusion constructs in a particular expression vector introduced into a particular host cell.

Recombinant Expression of Genes Encoding Endomannosidase

[0133] Another feature of the invention is the recombinant expression of the nucleic acid sequences encoding the endomannosidase. The nucleic acid sequences are operatively linked to an expression control sequence in an appropriate expression vector and transformed in an appropriate host cell (**Example 3**). A wide variety of suitable vectors readily available in the art are used to express the fusion constructs of the present invention in a variety of host cells. The vectors **pSH278, pSH279** and **pSH280 (Example 4)** are a select few examples described herein suitable for expression of endomannosidase activity in a lower eukarote, *Pichia pastoris*. It is to be understood that a wide variety of vectors suitable for expression of endomannosidase activity in a selected host cell are encompassed within the present invention.

[0134] In one aspect of the invention, a lower eukaryotic host cell producing glucosylated high mannose structures is modified by introduction and expression of the endomannosidase of the present invention. For example, a host cell *P. pastoris* **RDP25** (*och1 alg3*) producing hexose 6 is modified by introduction and expression of the endomannosidase of the present invention. The host cell of the present invention produces a modified glycan converting $GlcMan_5GlcNAc_2$ to $Man_4GlcNAc_2$. Accordingly, in one embodiment, a lower eukaryotic host cell expressing the endomannosidase of the present invention catalyzes the removal of a molecule comprising at least one glucose residue and a mannose residue.

[0135] The activity of the recombinant nucleic acid molecules encoding the endomannosidase of the invention are described herein. Varied expression levels are quantified by the conversion of a glucosylated glycan $GlcMan_5GlcNAc_2$ to a deglucosylated glycan $Man_4GlcNAc_2$. In one embodiment, the conversion of $GlcMan_5GlcNAc_2$ to $Man_4GlcNAc_2$ is partial (**Fig. 10, 11**).

[0136] In another embodiment, the conversion of $GlcMan_5GlcNAc_2$ to $Man_4GlcNAc_2$ is complete. In a preferred embodiment, at least 30% of $GlcMan_5GlcNAc_2$ is converted to $Man_4GlcNAc_2$. In a more preferred embodiment, at least 60% of $GlcMan_5GlcNAc_2$ is converted to $Man_4GlcNAc_2$. In an even more preferred embodiment, at least 90% of $GlcMan_5GlcNAc_2$ is converted to $Man_4GlcNAc_2$. Furthermore, it is contemplated that other glucose containing glycans are removed by the endomannosidase of the present invention. For example, the endomannosidase of the present invention further comprises the activity of truncating a glycan $Glc_{1-3}Man_{9-5}GlcNAc_2$ to $Man_{8-4}GlcNAc_2$.

[0137] Additionally, a gene encoding a catalytically active endomannosidase is expressed in a lower eukaryotic host cell (e.g. *Pichia pastoris*) modifying the glycosylation on a protein of interest. In one embodiment, the endomannosidase of the present invention modifies glucosylated N-linked oligosaccharides on a protein of interest. The resulting protein produces a more human-like glycoprotein. A lower eukaryotic host cell modified by the endomannosidase of the invention produces a $Man_{8-4}GlcNAc_2$ glycoform from a glucosylated glycoform on a protein of interest (**Fig. 2**). For example, a strain of *P. pastoris* modified by the endomannosidae of the invention produces a $Man_4GlcNAc_2$ glycoform and decreased moiety of the glucosylated hexose 6 glycoform on a protein of interest (**Fig. 10B**). Subsequent α1,2-mannosidase digestion of the $Man_4GlcNAc_2$ glycoform results in a trimannosyl core (**Fig. 10C**). Accordingly, the present invention provides a catalytically active endomannosidase in a lower eukaryotic host cell that converts a glucosylated glycoform to a desired glycoform on a therapeutic protein of interest.

[0138] Therapeutic proteins are typically administered by injection, orally, pulmonary, or other means. Examples of suitable target glycoproteins which may be produced according to the invention include, without limitation: erythropoietin, cytokines such as interferon-α, interferon-β, interferon-γ, interferon-ω, and granulocyte-CSF, coagulation factors such as factor VIII, factor IX, and human protein C, soluble IgE receptor α-chain, IgG, IgG fragments, IgM, interleukins, urokinase, chymase, and urea trypsin inhibitor, IGF-binding protein, epidermal growth factor, growth hormone-releasing factor, annexin V fusion protein,

angiostatin, vascular endothelial growth factor-2, myeloid progenitor inhibitory factor-1, osteoprotegerin, α-1-antitrypsin and α– feto proteins.

Promoters

[0139] In another aspect of the invention, the rat liver endomannosidase (Genbank gi:2642186), the human endomannosidase (Genbank gi:20547442) or the mouse mannosidase (Genbank AK030141) is cloned into a yeast integration plasmid under the control of a constitutive promoter to optimize the amount of endomannosidase activity while restricting adverse effects on the cell. This involves altering promoter strength and optionally includes using an inducible promoter to better control the expression of these proteins.

[0140] In addition to expressing the wild-type endomannosidase, modified forms of the endomannosidase are expressed to enhance cellular localization and activity. Varying lengths of the catalytic domain of endomannosidase is fused to endogenous yeast targeting regions as described in WO 02/00879. The catalytically active fragment encoding the endomannosidase genes are cloned into a yeast integration plasmid under the control of a constitutive promoter. This involves altering the promoter strength and may include using an inducible promoter to better control the expression of these proteins. Furthermore, to increase enzyme activity, the protein is mutated to generate new characteristics. The skilled artisan recognizes the routine modifications of the procedures disclosed herein may provide improved results in the production of unglucosylated glycoprotein of interest.

Codon Optimization

[0141] It is also contemplated that the nucleic acids of the present invention may be codon optimized resulting in one or more changes in the primary amino acid sequence, such as a conservative amino acid substitution, addition, deletion or combination thereof.

Secreted Endomannosidase

[0142] In another feature of the invention, a soluble secreted endomannosidase is expressed in a host cell. In a preferred embodiment, a soluble mouse or human endomannosidase is recombinantly expressed. A soluble endomannosidase lacks cellular localization signal that normally localizes to the Golgi apparatus or bind to the cell membrane. Expression of the catalytic domain of the endomannosidase to produce a soluble recombinant enzyme, which lacks the transmembrane domain, can be fused in-frame to a second domain or a tag that facilitates its purification. The secreted rat and human endomannosidase of the present invention from *P. pastoris* is shown in **Fig. 9** (Example 8).

[0143] Expressed endomannosidase is particularly useful for *in vitro* modification of glucosylated glycan structures. In a more preferred embodiment, the recombinant endomannosidase is used to produce unglucosylated glycan intermediates in large scale glycoprotein production. **Fig. 13** shows the activity of the rat (**Fig. 13B**) and human (**Fig. 13C**) endomannosidase that have cleaved the glucose-α1,3-mannose dimer on the glycan intermediate $GlcMan_5GlcNAc_2$ converting it to $Man_4GlcNAc_2$. (**See Fig. 14**). Accordingly, the endomannosidase of the present invention is used to modify glucosylated glycans *in vitro*. In addition, such soluble endomannosidase are purified according to methods well-known in the art.

[0144] The secreted endomannosidases converts glucosylated structures (e.g., $GlcMan_5GlcNAc_2$) **Fig. 14**(i) to deglucosylated structures (e.g., $Man_4GlcNAc_2$) **Fig. 14**(ii) by hydrolyzing at least one glucose residue and one mannose residue on an oligosaccharide. For example, a glucose-α1,3-mannose dimer is cleaved from the glucosylated oligosaccharide by the endomannosidase as shown in **Fig. 14**. Subsequent α1,2-mannosidase digestion **Fig. 14**(iii) results in the structure: $Man_3GlcNAc_2$ indicating an additional Manα1,2 on the trimannosyl core.

Host Cells

[0145] A number of host cells can be used to express the endomannosidase of the present invention. For example, the endomannosidase can be expressed in mammalian, plant, insect, fungal, yeast, algal or bacterial cells. For the

modification of glucosylation on a protein of interest, preferred host cells are lower eukaryotes producing $Glc_{1-3}Man_{9-5}GlcNAc_2$ structures. Additionally, other host cells producing a mixture of glucosylated glycans are selected. For example, a host cell (e.g., *P. pastoris* **RDP25**) producing the glucosylated structures such as $GlcMan_5GlcNAc_2$ in addition to unglucosylated structures such as $Man_6GlcNAc_2$ and its isomers is selected.

[0146] Preferably, a lower eukaryotic host cell is selected from the group consisting of *Pichia pastoris, Pichia finlandica, Pichia trehalophila, Pichia koclamae, Pichia membranaefaciens, Pichia opuntiae, Pichia thermotolerans, Pichia salictaria, Pichia guercuum, Pichia pijperi, Pichia stiptis, Pichia methanolica, Pichia sp., Saccharomyces cerevisiae, Saccharomyces sp., Hansenula polymorpha, Kluyveromyces sp., Kluyveromyces lactis, Candida albicans, Aspergillus nidulans, Aspergillus niger, Aspergillus oryzae, Trichoderma reesei, Chrysosporium lucknowense, Fusarium sp., Fusarium gramineum, Fusarium venenatum* and *Neurospora crassa*.

[0147] Other hosts may include well-known eukaryotic and prokaryotic hosts, such as strains of *E. coli*, *Pseudomonas*, *Bacillus*, *Streptomyces*, and animal cells, such as Chinese Hamster Ovary (CHO; e.g., the alpha-glucosidase I deficient strain Lec-23), R1.1, B-W and L-M cells, African Green Monkey kidney cells (e.g., COS 1, COS-7, BSC1, BSC40, and BMT10), insect cells (e.g., Sf9), and human cells (e.g., HepG2) and plant cells in culture.

Methods For Modifying Glucosylated N-Glycans

[0148] In another aspect of the invention, herein is provided a method for modifying the glucosylated glycans by introducing and expressing the endomannosidase of the present invention. **Fig. 1**, as highlighted, shows the endomannosidase cleavage of the mono-, di-, and tri-glucosylated glycans, represented by the second and third glucose residues. Accordingly, the endomannosidase enzyme of the present invention is introduced into the Golgi of host (e.g.yeast) to enhance the efficiency of deglucosylation, and thus enhancing subsequent trimming of the mannan structure prior to the addition of further sugars to produce a more human-like N-linked glycosylation structure (**Fig. 2**).

[0149] In a further aspect of the invention, introduction of the endomannosidase into the Golgi (e.g. yeast) provides a method of recovering glucosylated glycoproteins that have entered the Golgi and are thus no longer accessible to the ER glucosidase I and II enzymes. The endomannosidase of the present invention can process such glucosylated structures; for example, $Glc_{1-3}Man_{9-5}GlcNAc_2$ to $Man_{8-4}GlcNAc_2$, highlighted by the four mannose residues as shown in **Fig. 2**. Accordingly, the present invention provides a quality control mechanism wherein the recovered glucosylated oligosaccharides are deglucosylated.

[0150] Moreover, it is contemplated that the use of the endomannosidase obviates the need for the glucosidase I and II enzymes required in the early steps of glycan trimming. In one embodiment, a host cell of the present invention may be deficient in glucosidase I and/or II activity. In the absence of glucosidase I or II activities, a host cell of the present invention may still exhibit a glucose catalyzing activity through the endomannosidase. Accordingly, herein is provided a method of introducing a nucleic acid encoding an endomannosidase into a host (e.g. yeast), upon expression, modifies glucosylated glycoproteins that have entered the Golgi, which are are no longer accessible to the ER glucosidase I and glucosidase II enzymes. Preferably, the nucleic acid encoding the enzyme of the present invention cleaves a composition comprising at least one glucose residue and one mannose residue linked to an oligosaccharide (**Fig. 2**). More preferably, a Glcα1,3Man dimer, $Glc_2\alpha$1,3Man trimer or $Glc_3\alpha$1,3Man tetramer are cleaved according to the method of the present invention..

[0151] It will be a matter of routine experimentation for one skilled in the art to use the method described herein to optimize production of deglucosylated glycans (e.g. $Man_4GlcNAc_2$) using a selected fusion construct in a particular expression vector and host cell line. Accordingly, routine modifications can be made in the lower eukaryotic host cell expressing the endomannosidase of the present invention, which converts glucosylated glycans to deglucosylated glycans (e.g. $Man_4GlcNAc_2$) and subsequently to a desired intermediate for the production of therapeutic glycoproteins.

Introduction of Other Glycosylation Enzymes In Host Cells

[0152] Additionally, a set of modified glycosylation enzymes are introduced into host cells to enhance cellular localization and activity in producing glycoproteins of interest. This involves the fusion of varying lengths of the catalytic domains to yeast endogenous targeting regions as described in WO 02/00879. In one embodiment, a host cell *P. pastoris* **YSH97** (*och1 alg3* endmannosidase) is modified by introduction and expression of glycosylation enzymes or catalytically active fragment thereof selected from the group consisting of α1,2-mannosidase I and II, GnT I (*N*-acetylglucosaminyltransferase I), GnT II, GnT III, GnT IV, GnT V, GnT VI, galactosyltransferase, sialyltransferase and fucosyltransferase. Similarly, the enzymes' respective transporters and their substrates (e.g. UDP-GlcNAc, UDP-Gal, CMP-NANA) are introduced and expressed in the host cells. See WO 02/00879.

Endomannosidase pH optimum

[0153] In another aspect of the invention, the encoded endomannosidase has a pH optimum between about 5.0 and about 8.5, preferably between about 5.2 and about 7.2 and more preferably about 6.2. In another embodiment, the encoded enzyme is targeted to the endoplasmic reticulum, the Golgi apparatus or the transport vesicles between ER, Golgi or the trans Golgi network of the host organism, where it removes glucosylated structures present on oligosaccharides. **Fig. 15** shows a pH optimum profile of the human endomannosidase (SEQ ID NO:2) (Example 9).

[0154] The following are examples which illustrate the compositions and methods of this invention. These examples should not be construed as limiting: the examples are included for the purposes of illustration only.

EXAMPLE 1

Strains, culture conditions, and reagents

[0155] *Escherichia coli* strains TOP10 or DH5α were used for recombinant DNA work. Protein expression in yeast strains were carried out at room temperature in a 96-well plate format with buffered glycerol-complex medium

(BMGY) consisting of 1% yeast extract, 2% peptone, 100 mM potassium phosphate buffer, pH 6.0, 1.34% yeast nitrogen base, 4 X 10^{-5}% biotin, and 1% glycerol as a growth medium. The induction medium was buffered methanol-complex medium (BMMY) consisting of 1.5% methanol instead of glycerol in BMGY. Minimal medium is 1.4% yeast nitrogen base, 2% dextrose, 1.5% agar and 4 X 10^{-5} % biotin and amino acids supplemented as appropriate. Restriction and modification enzymes were from New England BioLabs (Beverly, MA). Oligonucleotides were obtained from the Dartmouth College Core facility (Hanover, NH) or Integrated DNA Technologies (Coralville, IA). MOPS, sodium cacodylate, manganese chloride were from Sigma (St. Louis, MO). Trifluoroacetic acid (TFA) was from Sigma/Aldrich, Saint Louis, MO. The enzymes N-glycosidase F, mannosidases, and oligosaccharides were obtained from Glyko (San Rafael, CA). DEAE ToyoPearl resin was from TosoHaas. Metal chelating "HisBind" resin was from Novagen (Madison, WI). 96-well lysate-clearing plates were from Promega (Madison, WI). Protein-binding 96-well plates were from Millipore (Bedford, MA). Salts and buffering agents were from Sigma (St. Louis, MO). MALDI matrices were from Aldrich (Milwaukee, WI).

EXAMPLE 2
Cloning of Human and Mouse Endomannosidases

[0156] As a positive control, we amplified the region homologous to the putative catalytic domain of the rat mannosidase gene using specific primers 5'-gaattcgccaccatggatttccaaaagagtgacagaatcaacag-3' (SEQ ID NO: 11) and 5'-gaattcccagaaacaggcagctggcgatc-3' (SEQ ID NO: 12) and subcloned the resultant region into a yeast integration plasmid using standard recombinant DNA techniques (*See, e.g.*, Sambrook et al. (1989) *Molecular Cloning, A Laboratory Manual* (2nd ed.), Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. and references cited therein, all incorporated reference; see also **Example 3**).

[0157] To identify the sequence of and isolate the ORF of the human endomannosidase, we performed a protein BLAST search using the rat endomannosidase protein sequence (Genbank gi:2642187) and identified a hypothetical human protein (Genbank gi:20547442) of 290 amino acids in length

which shows 88% identity and 94% similarity to amino acids 162 to 451 of the rat ORF (**Fig. 3A**). The DNA 5'-terminus of this human sequence was analyzed using translated BLAST and another hypothetical human protein (Genbank gi:18031878) was identified that possessed 95% identity over the first 22 amino acids of the search sequence but then terminates in a stop codon (**Fig. 3B**). Reading-frame analysis of this second sequence indicated that 172 amino acids were in-frame upstream of the homologus region (**Fig. 3C**). Combining both these 5' and 3' regions produced a putative sequence with an ORF of 462 amino acids (**Fig. 4**) and a predicted molecular mass of 54 kDa.

[0158] To confirm that the two human sequences are one entire ORF, we designed primers specific to the 5'-terminus of the gi:18031877 ORF and the 3'-terminus of the gi: 20547441 ORF (5'-atggcaaagtttcggagaaggacttgc-3' (SEQ ID NO: 13) and 5'- ttaagaaacaggcagctggcgatctaatgc-3' (SEQ ID NO: 14) respectively). These primers were used to amplify a 1389 bp fragment from human liver cDNA (Clontech, Palo Alto, CA) using Pfu Turbo DNA polymerase (Stratagene, La Jolla, CA) as recommended by the manufacturers, under the cycling conditions: 95°C for 1min, 1 cycle: 95°C for 30sec, 60 °C for 1min, 72 °C for 2.5min, 30 cycles; 72 °C for 5min, 1 cycle. The DNA fragment produced was incubated with Taq DNA polymerase for 10 min at 68 °C and TOPO cloned into pCR2.1 (Invitrogen, Carlsbad, CA). ABI DNA sequencing confirmed that both of the human sequences identified by BLAST searching produced one complete ORF, this confirmed construct was named **pSH131**.

[0159] The endomannosidase gene from mouse may be similarly amplified and isolated. (See also, e.g., Sambrook et al. (1989) *Molecular Cloning, A Laboratory Manual* (2nd ed.), Cold Spring Harbor Laboratory, Cold Spring Harbor, NY., Innis et al. (1990) *PCR Protocols: A Guide to Methods and Applications*, Academic Press, New York, NY and references cited therein, all incorporated reference). The primers 5'-atggcaaaatttcgaagaaggacctgcatc-3' mEndo forward (SEQ ID NO: 15) and 5'-ttatgaagcaggctgctgttgatccaatgc-3' mEndo reverse (SEQ ID NO: 16) are used to generate the mouse full-length endomannosidase open reading frame.

EXAMPLE 3
Generation of Recombinant Endomannosidase Constructs and Expression

[0160] To generate a yeast secreted form of the human endomannosidase, a region encoding the putative catalytic domain was expressed in the EasySelect *Pichia* Expression kit (Invitrogen) as recommended by the manufacturer. Briefly, PCR was used to amplify the ORF fragment from 178 to 1386 bases from pSH131 using the primers hEndo Δ59 forward and hEndo Δstop reverse (5'-gaattcgccaccatggatttccaaaagagtgacagaatcaacag-3' (SEQ ID NO: 11) and 5'-gaattcccagaaacaggcagctggcgatc-3' (SEQ ID NO: 12), respectively, with an *EcoRI* restriction site engineered into each). The conditions used with Pfu Turbo were: 95°C for 1 min, 1 cycle; 95°C for 30 sec, 55°C for 30 sec, 72°C for 3 min, 25 cycles; 72°C for 3 min, 1 cycle. The product was incubated with *Taq* DNA polymerase, TOPO cloned and ABI sequenced as described above. The resulting clone was designated **pSH178**. From this construct, the human endomannosidase fragment was excised by digestion with *Eco*RI and subcloned into pPicZαA (Invitrogen, Carlsbad, CA) digested with the same enzyme, producing pAW105. This construct was transformed into the *Pichia pastoris* yeast strain GS115 supplied with the EasySelect *Pichia* Expression kit (Invitrogen, Carlsbad, CA), producing the strain **YSH16**. Subsequently, the strain was grown in BMGY to an OD_{600} of 2 and induced in BMMY for 48 h at 30°C, as recommended by the kit manufacturers.

[0161] To confirm that the isolated ORF was an endomannosidase, the previously reported rat liver endomannosidase was amplified and expressed in parallel as a positive control. Briefly, the fragment encoding amino acids 49 to 451 of the rat endomannosidase, corresponding to the putative catalytic domain, was amplified from rat liver cDNA (Clontech) using the same conditions as described for the human endomannosidase above. The primers used were rEndo Δ48 forward and rEndo Δstop reverse (5'-gaattcgccaccatggacttccaaaggagtgatcgaatcgacatgg-3' (SEQ ID NO: 17) and 5'-gaattccctgaagcaggcagctgttgatcc-3' (SEQ ID NO: 18), respectively, with an EcoRI restriction site engineered into each). The PCR product was cloned into pCR2.1, sequenced and the resultant construct named **pSH179**. Subsequently, the rat

endomannosidase was subcloned into pPicZαA (Invitrogen, Carlsbad, CA) and expressed in GS115 (Invitrogen, Carlsbad, CA) as described above, producing **pAW106** and **YSH13**.

[0162] To N-terminal tag recombinant human and rat endomannosidases, a double FLAG tag was engineered 3' to the Kex2 cleavage site of the alpha mating factor and 5' to the EcoRI restriction used for endomannosidase cloning in pPicZαA, as follows. Briefly, the phosphorylated oligonucleotides FLAG tag forward and FLAG tag reverse (5'-P-aatttatggactacaaggatgacgacgacaagg-3' (SEQ ID NO: 19) and 5'-P-aattccttgtcgtcgtcatccttgtagtccata-3' (SEQ ID NO: 20)) were annealed as described in Sambrook et al. (1989), *supra*, and ligated into pPicZαA digested with *EcoRI* and dephosphorylated with calf alkaline phosphatase. A construct containing two tandem FLAG tags in the correct orientation was named **pSH241**. Subsequently, rat and human endomannosidases were digested from pSH179 and pSH178 with *EcoRI* and ligated into pSH241, digested with the same enzyme. The resultant rat and human endomannosidase constructs were named **pSH245** and **pSH246**, respectively. Transformation of these constructs into GS115 (Invitrogen, Carlsbad, CA) produced the strains **YSH89** and **YSH90**, respectively. Expression of endomannosidase activities in these strains was studied as described above.

EXAMPLE 4
Expression of rat endomannosidases in *P. pastoris*

[0163] The catalytic domain of rat endomannosidase was amplified from pSH179 using the primers rat Endomannosidase Δ48 *AscI* and rEndo *PacI* (5'-ggcgcgccgacttccaaaggagtgatcgaatcgacatgg-3' (SEQ ID NO: 21) and 5'-ccttaattaattatgaagcaggcagctgttgatccaatgc-3' (SEQ ID NO: 22), encoding *AscI* and *PacI* restriction sites respectively). These primers were used to amplify a 1212 bp fragment from pSH179 using Pfu Turbo DNA polymerase (Stratagene) as recommended by the manufacturers, under the cycling conditions: 95°C for 1 min, 1 cycle: 95°C for 30 sec, 60°C for 1 min, 72°C for 2.5 min, 30 cycles; 72°C for 5 min, 1 cycle. The DNA fragment produced was incubated with *Taq* DNA polymerase for 10 min at 68°C and TOPO cloned into pCR2.1 (Invitrogen, Carlsbad, CA). ABI DNA sequencing confirmed that both of the human sequences

identified by BLAST searching produced one complete ORF. This confirmed construct was named **pSH223**. Subsequently, the rat endomannosidase fragment was digested from this construct and ligated into the yeast expression vector pRCD259, giving the construct **pSH229**. The expression construct contains the hygromycin selection marker; GAPDH promoter and CYC1 terminator, with the cloning sites *NotI*, *AscI* and *PacI* located between these two regions; *URA3* targeting integration region; and a fragment of the pUC19 plasmid to facilitate bacterial replication.

EXAMPLE 5
Expression Vectors and Integration

[0164] To express the rat endomannosidase proteins in yeast, the cDNA encoding the catalytic domain was cloned into the expression vector pRCD259 producing the vector **pSH229** (See Example 4). Subsequently, cDNAs encoding Gls1(s), Van1(s) and Mnn11(m) leaders were cloned 5' to the cDNA encoding the rat endomannosidase catalytic domain producing the plasmids **pSH278** (rEndo Δ48 Gls1s leader), **pSH279** (rEndo Δ48 Van1s leader) and **pSH280** (rEndo Δ48 Mnn11m leader). Integration was confirmed by colony PCR with the resultant positive clones being analyzed to determine the N-glycan structure of a secreted reporter protein.

EXAMPLE 6
Northern Blot Analysis

[0165] Tissue distribution of human endomannosidase transcript was determined with a human Multiple Tissue Northern blot (Clontech) representing 2μg of purified poly A^+ RNA from each of the tissues according to the instructions of the manufacturer. The 547 bp human endomannosidase DNA probe (843-1389) used was generated using the RadPrime DNA Labeling System (Invitrogen, Carlsbad, CA) and [^{32}P]dCTP. The results are shown in **Fig. 8**.

EXAMPLE 7

SDS-PAGE and Western Blotting

[0166] Media from the *P. pastoris* cultures were analyzed for endomannosidase secretion by running samples on a 10% SDS-PAGE (Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature, 227, 680-685) using the Bio-Rad Mini-Protean II apparatus. The proteins were then transferred onto a nitrocellulose membrane (Schleicher & Schuell, Keene, NH). Recombinant endomannosidase was detected using the anti-FLAG M2 monoclonal antibody in combination with a goat anti-mouse HRP-conjugated secondary antibody and visualized with the ECL Western detection system (Amersham Biosciences) according to the manufacturer's instructions. Media from GS115 (Invitrogen, Carlsbad, CA) was used as a control. The results are shown in **Fig. 9**.

EXAMPLE 8

In vitro Characterization of Recombinant Endomannosidase

[0167] $GlcMan_5GlcNAc_2$, a substrate for endomannosidase assays, was isolated from the *och1 alg3* mutant strains **RDP25** (WO 03/056914A1) (Davidson et al, 2003 in preparation). 2-aminobenzamide-labeled $GlcMan_5GlcNAc_2$ was added to 10 µl of culture supernatant and incubated at 37^0C for 8 h or overnight. 10 µl of water was then added and subsequently the glycans were separated by size and charge using an Econosil NH_2 4.6 X 250 mm, 5 micron bead, amino-bound silica column (Altech, Avondale, PA) following the protocol of Choi et al , *Proc. Natl. Acad. Sci. U. S. A.* 100(9):5022-5027 (2003).

EXAMPLE 9

pH and Temperature Optima Assays of Engineered endo α-1,2-mannosidase

[0168] Fluorescence-labeled $GlcMan_5GlcNAc_2$ (0.5 µg) was added to 20µL of supernatant adjusted to various pH (Table 2) and incubated for 8 hours at room temperature. Following incubation the sample was analyzed by HPLC using an Econosil NH2 4.6 X 250 mm, 5 micron bead, amino-bound silica column (Altech, Avondale, PA). The flow rate was 1.0 ml/min for 40 min and the column was

maintained to 30°C. After eluting isocratically (68% A:32% B) for 3 min, a linear solvent gradient (68% A:32% B to 40% A:60% B) was employed over 27 min to elute the glycans (18). Solvent A (acetonitrile) and solvent B (ammonium formate, 50 mM, pH 4.5. The column was equilibrated with solvent (68% A:32% B) for 20 min between runs. The following table shows the amount (%) of $Man_4GlcNAc_2$ produced from $GlcMan_5GlcNAc_2$ at various pHs (**Fig. 15, Table 2**).

Table 2. pH Optimum of Human Endomannosidase

pH	% of Man4
4	0
4.5	0
5	4.5
5.5	29.6
6	51.4
6.5	52
7	41.3
7.5	30
8.5	20

[0169] The temperature optimum for human endomannosidase was similarly examined by incubating the enzyme substrate with culture supernatant at different temperatures (room temperature, 30°C and 37°C), 37°C being the optimum.

EXAMPLE 10

Reporter protein expression, purification and release of N-linked glycans

Protein Purification

[0170] Kringle 3 (K3) domain, under the control of the alcohol oxidase 1 (AOX1) promoter, was used as a model protein. Kringle 3 was purified using a 96-well format on a Beckman BioMek 2000 sample-handling robot (Beckman/Coulter Ranch Cucamonga, CA). Kringle 3 was purified from expression media using a C-terminal hexa-histidine tag (Choi et al. 2003, *supra*). The robotic purification is an adaptation of the protocol provided by Novagen for their HisBind resin. Briefly, a 150uL (μL) settled volume of resin is poured into the wells of a 96-well lysate-binding plate, washed with 3 volumes of water and charged with 5 volumes of 50mM NiSO4 and washed with 3 volumes of binding buffer (5mM imidazole, 0.5M NaCl, 20mM Tris-HCL pH7.9). The protein

expression media is diluted 3:2, media/PBS (60mM PO4, 16mM KCl, 822mM NaCl pH7.4) and loaded onto the columns. After draining, the columns are washed with 10 volumes of binding buffer and 6 volumes of wash buffer (30mM imidazole, 0.5M NaCl, 20mM Tris-HCl pH7.9) and the protein is eluted with 6 volumes of elution buffer (1M imidazole, 0.5M NaCl, 20mM Tris-HCl pH7.9). The eluted glycoproteins are evaporated to dryness by lyophilyzation.

Release of N-linked Glycans

[0171] The glycans are released and separated from the glycoproteins by a modification of a previously reported method (Papac et al., *Glycobiology* 8(5):445-54 (1998)). The wells of a 96-well MultiScreen IP (Immobilon-P membrane) plate (Millipore) were wetted with 100uL of methanol, washed with 3x150uL of water and 50uL of RCM buffer (8M urea, 360mM Tris, 3.2mM EDTA pH8.6), drained with gentle vacuum after each addition. The dried protein samples were dissolved in 30uL of RCM buffer and transferred to the wells containing 10uL of RCM buffer. The wells were drained and washed twice with RCM buffer. The proteins were reduced by addition of 60uL of 0.1M DTT in RCM buffer for 1hr at 37°C. The wells were washed three times with 300uL of water and carboxymethylated by addition of 60uL of 0.1M iodoacetic acid for 30min in the dark at room temperature. The wells were again washed three times with water and the membranes blocked by the addition of 100uL of 1% PVP 360 in water for 1hr at room temperature. The wells were drained and washed three times with 300uL of water and deglycosylated by the addition of 30uL of 10mM NH_4HCO_3 pH 8.3 containing one milliunit of N-glycanase (Glyko). After incubting for 16 hours at 37°C, the solution containing the glycans was removed by centrifugation and evaporated to dryness.

Miscellaneous: Proteins were separated by SDS/PAGE according to Laemmli (Laemmli 1970).

EXAMPLE 11
Matrix Assisted Laser Desorption Ionization Time of Flight Mass Spectrometry

[0172] Molecular weights of the glycans were determined using a Voyager DE PRO linear MALDI-TOF (Applied Biosciences) mass spectrometer using delayed

extraction. The dried glycans from each well were dissolved in 15uL of water and 0.5uL spotted on stainless steel sample plates and mixed with 0.5uL of S-DHB matrix (9mg/mL of dihydroxybenzoic acid, 1mg/mL of 5-methoxysalicilic acid in 1:1 water/acetonitrile 0.1% TFA) and allowed to dry.

[0173] Ions were generated by irradiation with a pulsed nitrogen laser (337nm) with a 4 ns pulse time. The instrument was operated in the delayed extraction mode with a 125 ns delay and an accelerating voltage of 20kV. The grid voltage was 93.00%, guide wire voltage was 0.10%, the internal pressure was less than 5 X 10-7 torr, and the low mass gate was 875Da. Spectra were generated from the sum of 100-200 laser pulses and acquired with a 2 GHz digitizer. $Man_5GlcNAc_2$ oligosaccharide was used as an external molecular weight standard. All spectra were generated with the instrument in the positive ion mode. The estimated mass accuracy of the spectra was 0.5%.

EXAMPLE 12
A Combinatorial Library To Produce a Chimeric Endomannosidase Protein

[0174] A library of human, mouse, rat and/or any combination of mixed endomannosidases characterized by catalytic domains having a range of temperature and pH optima is generated following published procedures (see, e.g., WO 02/00879; Choi et al. 2003, *supra* and the publication of United States Application No. 10/371,877 (filed Feb. 20, 2003)). This library will be useful for selecting one or more sequences which encode a protein having endomannosidase activity that performs optimally in modifying the glycosylation pattern of a reporter protein to produce a desired glycan structure when expressed in a lower eukaryotic host cell such as a yeast. It is expected to be advantageous to target the catalytic domain of the endomannosidase to a specific cellular compartment. The DNA combinatorial library approach (in-frame fusion between a targeting peptide and an enzymatic domain) enables one to identify a chimeric molecule which expresses an endomannosidase activity in a desired or an efficient way in the host cell used for the seletion. An endomannosidase sequence is expressed in a number of expression systems -- including bacterial, yeast and mammalian cells, to characterize the encoded protein.

[0175] To generate a human-like glycoform in a host, e.g., a microorganism, the host is engineered to express an endomannosidase enzyme (such as the human or mouse endomannosidase described herein) which hydrolyzes mono-, di- and tri-glucosylated high mannose glycoforms, removing the glucose residue(s) present and the juxta-positioned mannose (see **Fig. 1**). A DNA library comprising sequences encoding cis and medial Golgi localization signals (and optionally comprising ER localization signals) is fused in-frame to a library encoding one or more endomannosidase catalytic domains. The host organism is a strain, e.g. a yeast, that is deficient in hypermannosylation (e.g. an *och1* mutant) and preferably, provides *N*-glycans having the structure $GlcNAcMan_5GlcNAc_2$ in the Golgi and/or ER. (Endomannosidase can hydrolyze $Glc_{1-3}Man_{9-5}GlcNAc_2$ to $Man_{8-4}GlcNAc_2$, so the preferred $GlcNAcMan_5GlcNAc_2$ structure is not essential). After transformation, organisms having the desired glycosylation phenotype are selected. Preferably, the endomannosidase activity removes a composition comprising at least a glucose residue and one mannose residue on an oligosaccharide. An *in vitro* assay is used in one method. The desired structure is a substrate for the enzyme alpha 1,2-mannosidase (see **Fig. 2**). Accordingly, single colonies may be assayed using this enzyme *in vitro*

[0176] The foregoing *in vitro* assays are conveniently performed on individual colonies using high-throughput screening equipment. Alternatively, a lectin binding assay is used. In this case the reduced binding of lectins specific for terminal mannoses allows the selection of transformants having the desired phenotype. For example, *Galantus nivalis* lectin binds specifically to terminal α-1,3-mannose, the concentration of which is reduced in the presence of operatively expressed endomannosidase activity. In one suitable method, *G. nivalis* lectin attached to a solid agarose support (available from Sigma Chemical, St. Louis, MO) is used to deplete the transformed population of cells having high levels of terminal α-1,3-mannose.